

Optimizing Age Penalty in Time-Varying Networks with Markovian and Error-Prone Channel State

Yuchao Chen ^{1,2}, Haoyue Tang ^{1,2}, Jintao Wang ^{1,2,3,*} and Jian Song ^{1,2,3}

- ¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; cyc20@mails.tsinghua.edu.cn (Y.C.); tanghaoyue13@tsinghua.org.cn (H.T.); jsong@tsinghua.edu.cn (J.S.)
² Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China
³ Research Institute of Tsinghua University in Shenzhen, Shenzhen 518057, China
* Correspondence: wangjintao@tsinghua.edu.cn

Abstract: In this paper, we consider a scenario where the base station (BS) collects time-sensitive data from multiple sensors through time-varying and error-prone channels. We characterize the data freshness at the terminal end through a class of monotone increasing functions related to Age of information (AoI). Our goal is to design an optimal policy to minimize the average age penalty of all sensors in infinite horizon under bandwidth and power constraint. By formulating the scheduling problem into a constrained Markov decision process (CMDP), we reveal the threshold structure for the optimal policy and approximate the optimal decision by solving a truncated linear programming (LP). Finally, a bandwidth-truncated policy is proposed to satisfy both power and bandwidth constraint. Through theoretical analysis and numerical simulations, we prove the proposed policy is asymptotic optimal in the large sensor regime.

Keywords: age of information; cross-layer design; constrained Markov decision process



Citation: Chen, Y.; Tang, H.; Wang, J.; Song, J. Optimizing Age Penalty in Time-Varying Networks with Markovian and Error-Prone Channel State. *Entropy* **2020**, *23*, 91. <https://doi.org/10.3390/e23010091>

Received: 20 November 2020

Accepted: 8 January 2021

Published: 10 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The requirements for data freshness in numerous emerging applications are becoming stricter [1,2]. However, the limited resources and bandwidth, together with the fading and error-prone channel characteristics, prevent the control terminal from obtaining the newest information. Moreover, the traditional optimization goals like low delay and high throughput cannot fully characterize the requirement of data freshness. Therefore, it is necessary to introduce new metrics to capture data freshness in such systems and design strategies to optimize the system performance in the presence of resource and environment restrictions.

Recently, a popular metric, Age of information (AoI), has been proposed in [3] to measure the data freshness. Since then, the optimization of age performance under different systems has been a research hotspot. The simple point-to-point system model has been studied in [3–11]. When update packets are generated by external sources and are queued in a buffer before transmission, queuing theory can be used to analyze the performance of such system, see, e.g., in, [3–8]. In [3], it is shown that the optimum packet generation rate of a first-come-first-served (FCFS) system should achieve a trade-off between throughput and delay. In [8], dynamic pricing is used as an incentive to balance the AoI evolution and the monetary payments to the users. Other studies [9–11] consider the generate-at-will system without a queue. Energy constraints are studied in [10,11] to find the trade-off between the age performance and energy consumption. In [11], both offline and online heuristic policies are proposed to optimize the average AoI, which outperform the greedy approach.

Apart from the point-to-point systems, scheduling strategies in the multi-user networks are studied in [12–19]. Different scheduling policies are studied in [12] to minimize the average AoI performance through unreliable channels, and Maatouk et al. verify the

asymptotic optimality of Whittle's index policy by setting an upper bound on the maximum AoI [13]. When each user has a minimum throughput requirement, the authors of [14] propose the Drift-Plus-Penalty policy using Lyapunov analysis to minimize the average AoI performance. In [17], a slotted ALOHA algorithm is proposed to minimize the average AoI, whose performance is verified in the large system regime.

Notice that in many general scenarios, like remote estimation [20,21], the proper evaluation of data freshness is a function of AoI instead of AoI itself. Therefore, the metric of Cost of Update Delay (CoUD) in [7] and age penalty function in [9] have been proposed to measure data freshness in a general setting. However, those works focus on minimizing data freshness in a single-user network, and the multi-user model is rarely considered. To fill this gap, we study a scenario where the base station (BS) collects time-sensitive data from multiple sensors through time-varying channels. We generalize our previous work [22] by considering a more realistic time-varying channel with packet loss and a more general age penalty measurement to model different application scenarios. The main contributions of the paper are listed as follows.

- We study the scheduling strategy for age penalty minimization in multi-sensor bandwidth constrained networks through time-varying and error-prone channel links with power limited sensors. To study a practical network, we model the channel to be a finite-state ergodic Markov chain. The packet loss probability and power consumption depend on the current channel state. Unlike previous work, we model the effect of data staleness in different scenarios via a class of monotone increasing function related to AoI.
- Through relaxing the hard bandwidth constraint and Lagrange multipliers, we decouple the multi-sensor optimization problem into several single-sensor constrained Markov decision process (CMDP) problems. To deal with the potential infinite age penalty, we deduce the threshold structure of the optimal policy and then obtain the approximate optimal single-sensor scheduling decision by solving a truncated linear programming (LP). We prove the solution to the LP is asymptotic optimal when the truncated threshold is sufficiently large.
- The sub-gradient ascend method is applied to find the optimal Lagrange multiplier to satisfy the relaxed bandwidth constraint. Finally, we propose the truncated stationary policy to meet the hard bandwidth constraint. The average performance of the strategy is verified through theoretical analysis and numerical simulations.

The remainder of this paper is organized as follows. The network model, the AoI metric, and the age penalty function are introduced in Section 2. In Section 3, we formulate the primal scheduling problem, and then decouple it into independent single-sensor problems through bandwidth relaxation and Lagrange multipliers. The approximate optimal policy for single-sensor problem is obtained in Section 4 by solving an LP. In Section 5, the asymptotic optimal truncated policy is proposed. Finally, Section 6 provides simulation results to verify the performance of the proposed truncated policy, and Section 7 draws the conclusion.

Notations: All the vectors and matrices are denoted in boldface. The probability of \mathcal{A} given \mathcal{B} is denoted as $\Pr(\mathcal{A}|\mathcal{B})$. Let $\mathbb{E}_\pi[X]$ be the expectation of variable X given π . The cardinality of a set Ω is written as $|\Omega|$.

2. System Model

2.1. Network Model

In this work, we consider a BS collecting update packets from N time-sensitive sensors through time-varying channels. The time is slotted, and $t \in \{1, 2, \dots, T\}$ is used to denote the current slot index. Define $u_n(t)$ to be the scheduling decision for sensor n in slot t , where $u_n(t) = 1$ means the sensor n chooses to send the newest packet, while $u_n(t) = 0$ means idling the channel link. Assume all the scheduling behaviors take place at the beginning of each slot and the packet transmission delay through all the channel links is one slot. Due to the limited bandwidth capacity of the BS, the total number of sensors to

be scheduled in one slot cannot be larger than M . Here, we assume $M < N$ so that the problem is nontrivial:

$$\sum_{n=1}^N u_n(t) \leq M, \forall t. \tag{1}$$

To model the time varying effect, we assume that the channel link connecting the BS and each sensor n is an ergodic Q -state Markov chain. Denote $q_n(t) \in \{1, 2, \dots, Q\}$ to be the channel state of link n in slot t . Without loss of generality, we assume that the channel state becomes more noisy as the state index becomes larger. Denote $\mathbf{P}_n = \{p_{ij}^n\}$ to be the Markov state transition matrix of link n , and the entry p_{ij}^n means the probability of changing into state j in the next slot given the current state i , i.e.,

$$p_{ij}^n \triangleq \Pr\{q_n(t+1) = j | q_n(t) = i\}. \tag{2}$$

Due to different channel states, the sensors should use different energy for both saving energy and successful decoding of the packet at the receiver. We denote $w(q)$ to be the energy consumption for scheduling when the channel state is q . Notice that the energy consumption tends to be larger as the channel state becomes more noisy to combat the channel fading, i.e., $w(1) < w(2) < \dots < w(Q)$. Besides, due to the power limit of each sensor n , the total average power consumption cannot exceed the upper bound, denoted by \mathcal{E}_n , i.e.,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \forall n, \tag{3}$$

where π is a scheduling policy.

Given channel state q , we assume that there exists the probability of packet loss $\varepsilon_{n,q}$ through link n due to decoding error or inaccurate estimation.

2.2. Age of Information and Age Penalty

In the network described above, the BS wishes to obtain the freshest information for further process or accurate prediction. We model the data staleness at the terminal end as a monotone increasing age penalty function $f(\cdot)$ of Age of information (AoI). By definition, AoI is the difference between the current time slot and the time slot that the freshest data received by the BS is generated by the sensor [3].

Let $x_n(t)$ be the AoI of sensor n in slot t . According to the definition, if the sensor is scheduled in slot t and there is no packet loss, then $x_n(t+1) = 1$; otherwise, $x_n(t+1) = x_n(t) + 1$. In conclusion, the AoI evolves as follows,

$$\Pr(x_n(t+1) = x' | x_n(t) = x, u_n(t) = u) = \begin{cases} 1 - \varepsilon_{n,q_n(t)}, & x' = 1, u = 1; \\ \varepsilon_{n,q_n(t)}, & x' = x + 1, u = 1; \\ 1, & x' = x + 1, u = 0; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

3. Problem Formulation and Decomposition

3.1. Problem Formulation

For given network, we measure the data freshness at the terminal side by computing the average age penalty under scheduling policy π , denoted by $J(\pi)$, i.e.,

$$J(\pi) = \lim_{T \rightarrow \infty} \frac{1}{NT} \mathbb{E}_\pi \left[\sum_{t=1}^T \sum_{n=1}^N f(x_n(t)) | \mathbf{x}(0) \right], \tag{5}$$

where $\mathbf{x}(0) = [x_1(0), x_2(0), \dots, x_N(0)]$ states the initial AoI of the system. In this work, we assume that the system is synchronized with all the sensors at the beginning, i.e., $x_n(0) = 1, \forall n$, and thus omit $\mathbf{x}(0)$ in the further analysis.

We denote Π_{CP} to be the set of all possible causal policies whose decisions are only based on current and historic information while satisfying both bandwidth and power constraints. Then, our goal is to optimize Equation (5) by choosing a scheduling policy $\pi \in \Pi_{CP}$. Therefore, the primal optimization problem can be written as

Problem 1 (Primal Scheduling Problem).

$$\min_{\pi \in \Pi_{CP}} \lim_{T \rightarrow \infty} \frac{1}{NT} \mathbb{E}_{\pi} \left[\sum_{n=1}^N \sum_{t=1}^T f(x_n(t)) \right], \quad (6a)$$

$$\text{s.t.} \quad \sum_{n=1}^N u_n(t) \leq M, \quad \forall t, \quad (6b)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \quad \forall n. \quad (6c)$$

3.2. Problem Decomposition

Notice that Equation (6b) is an integer programming where the exponential growth rate of state and action space set obstacles in solving Problem 1. Therefore, we formulate a relaxed version of Problem 1, where the primal bandwidth constraint in every slot is replaced by a time-average bandwidth constraint. We will then show that the relaxed problem can be solved by sensor level decoupling, which greatly reduces the cardinality of the state and action space.

Problem 2 (Relaxed Primal Scheduling Problem).

$$Age_R^* = \min_{\pi \in \Pi_{CP}} \lim_{T \rightarrow \infty} \frac{1}{NT} \mathbb{E}_{\pi} \left[\sum_{n=1}^N \sum_{t=1}^T f(x_n(t)) \right], \quad (7a)$$

$$\text{s.t.} \quad \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T \sum_{n=1}^N u_n(t) \right] \leq M, \quad (7b)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \quad \forall n. \quad (7c)$$

Denote π_R^* to be the optimal policy of Problem 2. The following theorem ensures that the optimal policy of the relaxed problem is composed of several local optimal policies π_n^* , each of which depends on its own channel state and AoI evolution regardless of others.

Theorem 1. *The optimal policy of Problem 2 can be decoupled into local optimal policies, i.e., $\pi_R^* = \pi_1^* \otimes \pi_2^* \otimes \dots \otimes \pi_N^*$. Each of the local policy π_n^* has the following properties.*

$$Age_R^* = \lim_{T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T \sum_{n=1}^N \mathbb{E}_{\pi_n^*} [f(x_n(t))],$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \mathbb{E}_{\pi_n^*} [u_n(t)] \leq M,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi_n^*} \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \quad \forall n.$$

The proof of Theorem 1 is provided in Appendix A.

In order to find the local optimal policies, we introduce a Lagrange multiplier $\lambda \geq 0$ to eliminate the relaxed bandwidth constraint, and the Lagrange function is as follows,

$$\mathcal{L}(\pi, \lambda) = \lim_{T \rightarrow \infty} \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T \mathbb{E}_{\pi_n} \left[f(x_n(t)) + \lambda u_n(t) - \lambda \frac{M}{N} \right],$$

where the Lagrange multiplier λ can be seen as a scheduling penalty which will increase the function value if there are more than M sensors to be scheduled per slot in average.

For fixed λ , we can now further decouple the relaxed scheduling problem into N single-sensor cross-layer designs, each of which has the corresponding power constraint in Equation (7c):

Problem 3 (Single-Sensor Decoupled Problem).

$$\min_{\pi_n \in \Pi_{CP}} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi_n} \left[\sum_{t=1}^T (f(x_n(t)) + \lambda u_n(t)) \right], \tag{8a}$$

$$\text{s.t. } \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi_n} \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n. \tag{8b}$$

As the resolution of each decoupled problem is independent of n , we omit the subscript n in further analysis.

4. Single-Sensor Problem Resolution

4.1. Constrained Markov Decision Process Formulation

First, we notice that the decoupled problem is a constrained Markov decision process of which the elements $(\mathbb{S}, \mathbb{A}, \Pr(\cdot|\cdot), c(\cdot))$ and constraint are explained as follows.

- **State Space:** The state of each sensor consists of two parts: the current AoI $x(t)$ and channel state $q(t)$. Thus, $\mathbb{S} = \{x \times q\}$ is infinite but countable.
- **Action Space:** There are two possible actions in the action space \mathbb{A} for the scheduling policy, denoted by $u(t)$. Action $u(t) = 1$ means the sensor chooses to schedule while $u(t) = 0$ means idling. Notice that here $u(t)$ does not need to satisfy the bandwidth constraint.
- **Probability Transition Function:** According to Equations (2) and (4), the probability transition function can be written out as follows.

$$\begin{aligned} \Pr((x + 1, q')|(x, q), u = 0) &= p_{qq'}, \\ \Pr((x + 1, q')|(x, q), u = 1) &= p_{qq'} \varepsilon_{n,q}, \\ \Pr((1, q')|(x, q), u = 1) &= p_{qq'} (1 - \varepsilon_{n,q}). \end{aligned}$$

- **One-Step Cost:** The one-step cost consists of two parts: the age penalty growth and scheduling penalty, which can be computed by

$$c_x(x(t), q(t), u(t)) = f(x(t)) + \lambda u(t). \tag{9}$$

And the one-step power consumption is

$$c_E(x(t), q(t), u(t)) = u(t) w(q(t)). \tag{10}$$

Now our goal is to optimize the following average one-step cost,

$$\min_{\pi \in \Pi_{CP}} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T c_x(x(t), q(t), u(t)) \right], \tag{11}$$

under the following average power constraint,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T c_E(x(t), q(t), u(t)) \right] \leq \mathcal{E}. \tag{12}$$

4.2. Characterization of the Optimal Policy

To search for the stationary optimal policy, we can further introduce another Lagrange multiplier $\nu \geq 0$ to eliminate the power constraint, i.e.,

$$\frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=1}^T c_x(x(t), q(t), u(t)) + \nu c_E(x(t), q(t), u(t)) \right] - \nu \mathcal{E}.$$

The multiplier ν can be viewed as a power penalty, which will increase the Lagrange function once the average power consumption exceeds the constraint. Then minimizing the above Lagrange function for fixed ν becomes an MDP problem without any constraint.

The following lemma ensures that the optimal stationary policy for the MDP problem has a threshold structure.

Lemma 1. *The optimal stationary policy of the unconstrained MDP problem has the threshold structure, i.e., given state (x, q) , there exists a threshold τ_q such that if $x \geq \tau_q$, then it is optimal to schedule the sensor; otherwise, idling is the optimal action.*

Proof sketch: The complete proof is provided in Appendices B and C, which is similar to Theorem 2 in [23]. Despite the complex proof, the intuition is simple. As it is optimal to schedule the sensor in state (x, q) , then it is also the optimal action in state (x', q) , $\forall x' > x$ because the AoI is much bigger.

4.3. Linear Programming Approximation

Now, we focus on finding the optimal stationary policy. Denote $\zeta_{x,q}$ to be the scheduling probability given state (x, q) and our goal is to find $\{\zeta_{x,q}^*\}$ that minimizes the objective function. In this part, we will approximate $\{\zeta_{x,q}^*\}$ by solving a truncated LP.

According to Lemma 1, for the optimal stationary policy, we can set a threshold $X > \max_q \tau_q$ and then we have $\zeta_{x,q}^* = 1, \forall x \geq X$. Next, we focus on searching for policies that possesses this threshold property as other policies are far from optimality.

Let $\mu_{x,q}$ be the steady distribution of state (x, q) . Then, define a new variable $y_{x,q} \triangleq \mu_{x,q} \zeta_{x,q}$. The following theorem converts the CMDP problem into an infinite LP problem.

Theorem 2. *The single-sensor decoupled problem is equal to the following infinite LP problem.*

$$\{\mu_{x,q}^*, y_{x,q}^*\} = \arg \min_{\mu_{x,q}, y_{x,q}} \sum_{q=1}^Q \sum_{x=1}^{\infty} (f(x) \mu_{x,q} + \lambda y_{x,q}), \tag{13a}$$

$$s.t. \sum_{q=1}^Q \sum_{x=1}^{\infty} \mu_{x,q} = 1, \tag{13b}$$

$$\mu_{1,q'} = \sum_{q=1}^Q \sum_{x=1}^{\infty} (1 - \varepsilon_q) p_{qq'} y_{x,q}, \tag{13c}$$

$$\mu_{x,q'} = \sum_{q=1}^Q p_{qq'} [\mu_{x-1,q} - (1 - \varepsilon_q) y_{x-1,q}], \quad \forall x \geq 2, \tag{13d}$$

$$\sum_{q=1}^Q \sum_{x=1}^{\infty} y_{x,q} w(q) \leq \mathcal{E}, \tag{13e}$$

$$0 \leq \mu_{x,q} \leq 1, \quad 0 \leq y_{x,q} \leq \mu_{x,q}. \tag{13f}$$

Proof. Let us consider the average cost of Equation (8a) by using $\mu_{x,q}$ and $y_{x,q}$. Invoking Equation (9), the one step cost of state (x, q) is either $f(x) + \lambda$ when scheduling or $f(x)$ when idling. Therefore, the time average cost can be computed as follows,

$$\sum_{q=1}^Q \sum_{x=1}^{\infty} \mu_{x,q} [\xi_{x,q} (f(x) + \lambda) + (1 - \xi_{x,q}) f(x)] = \sum_{q=1}^Q \sum_{x=1}^{\infty} (f(x) \mu_{x,q} + \lambda y_{x,q}),$$

which is equivalent to Equation (13a).

Similarly, according to Equation (10), the time average power consumption is

$$\sum_{q=1}^Q \sum_{x=1}^{\infty} \mu_{x,q} [\xi_{x,q} w(q)] = \sum_{q=1}^Q \sum_{x=1}^{\infty} y_{x,q} w(q),$$

which is exactly the LHS of Equation (13e).

Considering the property of steady probability distribution, Equation 13b,f are verified.

Finally, notice that the evolution of state (x, q) forms a Markov chain as depicted in Figure 1 (top) for $Q = 2$ as an example. We use $\alpha_{q,q'}^x = \Pr((x + 1, q') | (x, q))$ and $\beta_{q,q'}^x = \Pr((1, q') | (x, q))$ to denote the transition probability between the states, which can be computed as follows,

$$\alpha_{q,q'}^x = \begin{cases} (1 - \xi_{x,q}) p_{qq'} + \xi_{x,q} \varepsilon_q p_{qq'}, & x < X; \\ \varepsilon_q p_{qq'}, & x \geq X. \end{cases} \tag{14}$$

$$\beta_{q,q'}^x = \begin{cases} \xi_{x,q} (1 - \varepsilon_q) p_{qq'}, & x < X; \\ (1 - \varepsilon_q) p_{qq'}, & x \geq X. \end{cases} \tag{15}$$

According to the property of steady distribution, $\mu_{x,q}$ equals to the sum of the steady distribution which can be transferred to $\mu_{x,q}$ in the next slot times their transition probability. As depicted in Figure 1, $\mu_{2,2} = \mu_{1,1} \alpha_{1,2}^1 + \mu_{1,2} \alpha_{2,2}^1$ (see the dashed lines). Therefore, we can compute $\mu_{x,q}$ as follows,

$$\mu_{x,q'} = \begin{cases} \sum_{q=1}^Q \sum_{x'=1}^{\infty} \beta_{q,q'}^{x'} \mu_{x',q}, & x = 1; \\ \sum_{q=1}^Q \alpha_{q,q'}^{x-1} \mu_{x-1,q}, & x \geq 2, \end{cases}$$

which is equivalent to Equation 13c,d. \square

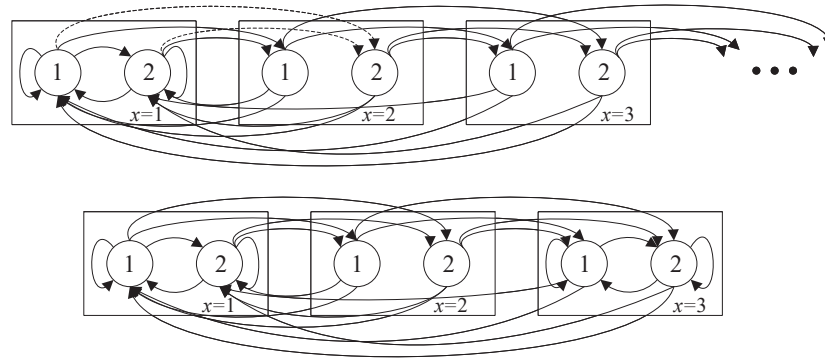


Figure 1. Illustration of the state transition graph for $Q = 2$ channel states without (top) and with (bottom) AoI truncation with AoI threshold $X = 3$. The numbers in circles are channel state index q , and the number in rectangles are AoI index x .

As the steady distribution is infinite, it is difficult to solve the problem exactly. Therefore, we approximate the optimization problem in Theorem 2 into a finite LP problem through truncation:

Problem 4 (Linear Programming Approximation).

$$\{\tilde{\mu}_{x,q}^*, \tilde{y}_{x,q}^*\} = \arg \min_{\mu_{x,q}, y_{x,q}} \sum_{q=1}^Q \sum_{x=1}^X (f(x)\mu_{x,q} + \lambda y_{x,q}), \tag{16a}$$

$$s.t. \sum_{q=1}^Q \sum_{x=1}^X \mu_{x,q} = 1, \tag{16b}$$

$$\mu_{1,q'} = \sum_{q=1}^Q \sum_{x=1}^X (1 - \varepsilon_q) p_{qq'} y_{x,q}, \tag{16c}$$

$$\mu_{x,q'} = \sum_{q=1}^Q p_{qq'} [\mu_{x-1,q} - (1 - \varepsilon_q) y_{x-1,q}], \forall 2 \leq x \leq X - 1, \tag{16d}$$

$$\mu_{X,q'} = \sum_{q=1}^Q p_{qq'} [\mu_{X-1,q} + \mu_{X,q} - (1 - \varepsilon_q)(y_{X-1,q} + y_{X,q})], \tag{16e}$$

$$\sum_{q=1}^Q \sum_{x=1}^X y_{x,q} w(q) \leq \mathcal{E}, \tag{16f}$$

$$0 \leq \mu_{x,q} \leq 1, 0 \leq y_{x,q} \leq \mu_{x,q}, \tag{16g}$$

$$y_{X,q} = \mu_{X,q}. \tag{16h}$$

After truncation, the optimal value of Equation (16a) is the lower bound of the objective function of the decoupled problem Equation (8a). The detailed proof is provided in Appendix D. The key concept is to set a threshold X and convert the Markov chain into a finite-state one (see Figure 1).

Moreover, the following theorem guarantees the lower bound obtained by the above LP problem is tight when X is sufficiently large. Thus, the approximate optimal solution $\{\tilde{\mu}_{x,q}^*, \tilde{y}_{x,q}^*\}$ performs close to the exact optimal solution $\{\mu_{x,q}^*, y_{x,q}^*\}$. Before displaying the theorem, first denote π_X^* and π^* to be the scheduling policy according to the approximate optimal solution $\{\tilde{\mu}_{x,q}^*, \tilde{y}_{x,q}^*\}$ by setting threshold X and optimal one $\{\mu_{x,q}^*, y_{x,q}^*\}$, respectively. Define $J_\infty(\pi)$ to be the age and scheduling penalty of the primal problem under policy π

and $J_X(\pi)$ is the approximate penalty when we set the age penalty $f(x) = f(X), \forall x \geq X$. Then, according to Equations (13a) and (16a), we have

$$J_\infty(\pi^*) = \sum_{q=1}^Q \sum_{x=1}^\infty (f(x)\mu_{x,q}^* + \lambda y_{x,q}^*), \quad J_X(\pi_X^*) = \sum_{q=1}^Q \sum_{x=1}^X (f(x)\tilde{\mu}_{x,q}^* + \lambda \tilde{y}_{x,q}^*).$$

Theorem 3. Assume $f(x)$ satisfies the following property: $\exists \epsilon \in [\sqrt{\epsilon_{max}}, 1)$ and constant $k > 0$ such that $f(X) = k\epsilon^{-X}$ and $f(x) \leq k\epsilon^{-x}, \forall x \geq X$, where $\epsilon_{max} = \max_q \epsilon_q$. Then, we have the following property,

$$J_\infty(\pi^*) - J_X(\pi_X^*) \leq \frac{k\epsilon^{\frac{X+1}{2} - \tau_{max}}}{1 - \epsilon_{max}}, \tag{17}$$

where $\tau_{max} = \max_q \tau_q$. As we see the above inequality, the difference between optimal solution of Theorem 2 and Problem 4 converges to 0 as the threshold X becomes sufficiently large.

The entire proof is provided in Appendix E.

After solving the above LP problem, we can obtain the approximate optimal scheduling probability $\{\tilde{\zeta}_{x,q}^*\}$ by setting a sufficiently large X and computing $\{\tilde{\mu}_{x,q}^*\}$ and $\{\tilde{y}_{x,q}^*\}$. Moreover, analogical to the threshold structure described in Lemma 1, $\{\tilde{\zeta}_{x,q}^*\}$ also has the following property.

Lemma 2. For any state (x, q) of each sensor, the optimal scheduling probability $\{\tilde{\zeta}_{x,q}^*\}$ is non-decreasing with x , i.e.,

$$\tilde{\zeta}_{x_1,q}^* \leq \tilde{\zeta}_{x_2,q}^*, \quad \forall x_1 \leq x_2.$$

The proof technique is similar to Lemma 1, so it is omitted here.

5. Multi-Sensor Problem Resolution

By now, through relaxing, decoupling and truncation, we have obtained the approximate solution to the single-sensor decoupled problem for fixed scheduling penalty λ . In this section, we will go back to solve the multi-sensor problem, and propose a truncated policy to meet the hard bandwidth constraint in Equation (6b).

5.1. The Relaxed Problem Resolution

First, we should choose the optimal λ such that the relaxed bandwidth constraint can be fully leveraged. Denote $g(\lambda) = \min_\pi L(\pi, \lambda)$ to be the Lagrange dual function, where we choose the approximate optimal policy $\pi^*(\lambda)$ by solving LP. Then, the dual function can be computed as follows,

$$g(\lambda) = \frac{1}{N} \left(\sum_{n=1}^N g_n(\lambda) - \lambda M \right),$$

where $g_n(\lambda) = \min_\pi L_n(\pi, \lambda)$ is the decoupled dual function for sensor n . According to the LP approximation, $g_n(\lambda)$ can be further written out as follows,

$$g_n(\lambda) = X_n(\lambda) + \lambda U_n(\lambda),$$

where $X_n(\lambda)$ is the average age penalty bounded by $\sum_{x=1}^X \sum_{q=1}^Q f(x)\tilde{\mu}_{x,q}^{n,*}$, and $U_n(\lambda)$ is the average scheduling probability, which equals to $\sum_{x=1}^X \tilde{y}_{x,q}^{n,*}$.

According to the work in [24], the optimal Lagrange multiplier λ^* satisfies

$$\lambda^* = \sup \left\{ \lambda \mid \sum_{n=1}^N U_n(\lambda) \leq M \right\}.$$

If $U_n(\lambda^*) = M$, then the optimal policy is just $\pi(\lambda^*)$. Otherwise, the optimal policy is a mixture of two policies, denoted by π_l and π_u , which can be computed by

$$\pi_l = \lim_{\lambda \rightarrow \lambda^{*-}} \pi(\lambda), \pi_u = \lim_{\lambda \rightarrow \lambda^{*+}} \pi(\lambda).$$

Then, we apply the sub-gradient ascend method to find the optimal solution, where the sub-gradient can be computed as follows,

$$d_\lambda g(\lambda) = \sum_{n=1}^N U_n(\lambda) - M = U(\lambda) - M, \tag{18}$$

where $U(\lambda) = \sum_{n=1}^N U_n(\lambda)$ is the total scheduling probability.

Choose $\lambda_0 = 0$ as the starting point, and compute the average scheduling probability $U(\lambda_0)$. If $U(\lambda_0) < M$, then it does not need to consider the bandwidth constraint, and thus the optimal solution has already been solved. Moreover, this optimal solution can also be viewed as the lower bound of the primal optimization problem, i.e.,

$$LB = \frac{1}{N} \sum_{n=1}^N \sum_{x=1}^X \sum_{q=1}^Q f(x) \tilde{\mu}_{x,q}^{n,*}(\lambda_0).$$

Otherwise, we need to increase the scheduling penalty through iterations. The update operation in iteration k can be written out as follows,

$$\lambda_{k+1} = \lambda_k + t_{k+1} d_\lambda g(\lambda_k),$$

where t_{k+1} is the step size in iteration k .

Moreover, the step size is determined as follows,

$$t_{k+1} = \begin{cases} \gamma t_k, & d_\lambda g(\lambda_k) d_\lambda g(\lambda_{k-1}) < 0; \\ t_k, & \text{otherwise,} \end{cases}$$

where $\gamma \in (0, 1)$ is a constant.

The determination of the step size above guarantees the algorithm converges from both sides. Therefore, after running the whole algorithm, we can obtain two different scheduling probabilities M_l and M_u :

$$M_l = \max_k U(\lambda_k), \text{ s.t. } d_\lambda g(\lambda_k) \leq 0; \tag{19}$$

$$M_u = \min_k U(\lambda_k), \text{ s.t. } d_\lambda g(\lambda_k) \geq 0. \tag{20}$$

Their corresponding optimal policies are denoted as $\{\tilde{\mu}_{x,q}^{n,l}, \tilde{y}_{x,q}^{n,l}\}$ and $\{\tilde{\mu}_{x,q}^{n,u}, \tilde{y}_{x,q}^{n,u}\}$. Then, the optimal stationary policy can be obtained by mixing these two policies:

$$\{\tilde{\mu}_{x,q}^{n,*}, \tilde{y}_{x,q}^{n,*}\} = \theta \{\tilde{\mu}_{x,q}^{n,u}, \tilde{y}_{x,q}^{n,u}\} + (1 - \theta) \{\tilde{\mu}_{x,q}^{n,l}, \tilde{y}_{x,q}^{n,l}\},$$

where the mixed coefficient can be computed as follows,

$$\theta = \frac{M - M_l}{M_u - M_l}.$$

Now, we have obtained the optimal stationary policy of the relaxed scheduling problem. The algorithm flow chart is listed in Algorithm 1. Once we obtain $\{\tilde{\mu}_{x,q}^{n,*}, \tilde{y}_{x,q}^{n,*}\}$, the optimal scheduling probability $\{\tilde{\xi}_{x,q}^{n,*}\}$ can be computed as follows,

$$\tilde{\zeta}_{x,q}^{n,*} = \begin{cases} 1, & \text{if } x > X \text{ or } \tilde{\mu}_{x,q}^{n,*} = 0 \text{ or } \tilde{\zeta}_{x-1,q}^{n,*} = 1; \\ \frac{\tilde{y}_{x,q}^{n,*}}{\tilde{\mu}_{x,q}^{n,*}}, & \text{otherwise.} \end{cases}$$

Algorithm 1 Construction of the optimal stationary policy

Initialization: $\lambda_{-1} = \lambda_0 = 0, \epsilon, t_0, \gamma, M_u$ and M_l
for each $n \in [1, N]$ **do**
 compute $\{\tilde{\mu}_{x,q}^{n,*}(\lambda_0), \tilde{y}_{x,q}^{n,*}(\lambda_0)\}$ and $U_n(\lambda_0)$
end for
if $d_{\lambda}g(\lambda_0) \leq 0$ **then**
 $\{\tilde{\mu}_{x,q}^{n,*}, \tilde{y}_{x,q}^{n,*}\} = \{\tilde{\mu}_{x,q}^{n,*}(\lambda_0), \tilde{y}_{x,q}^{n,*}(\lambda_0)\}$
else
 $k = 0$
 while $|\lambda_k - \lambda_{k-1}| > \epsilon$ or $d_{\lambda}g(\lambda_k) > 0$ **do**
 if $d_{\lambda}g(\lambda_k)d_{\lambda}g(\lambda_{k-1}) < 0$ **then**
 $t_{k+1} = \gamma t_k$
 else
 $t_{k+1} = t_k$
 end if
 $\lambda_{k+1} = \lambda_k + t_{k+1}d_{\lambda}g(\lambda_k)$
 for each $n \in [1, N]$ **do**
 compute $\{\tilde{\mu}_{x,q}^{n,*}(\lambda_{k+1}), \tilde{y}_{x,q}^{n,*}(\lambda_{k+1})\}$ and $U_n(\lambda_{k+1})$
 end for
 if $d_{\lambda}g(\lambda_{k+1}) \geq 0$ and $U(\lambda_{k+1}) \leq M_u$ **then**
 $\{\tilde{\mu}_{x,q}^{n,u}, \tilde{y}_{x,q}^{n,u}\} = \{\tilde{\mu}_{x,q}^{n,*}(\lambda_{k+1}), \tilde{y}_{x,q}^{n,*}(\lambda_{k+1})\}$
 $M_u = U(\lambda_{k+1})$
 end if
 if $d_{\lambda}g(\lambda_{k+1}) \leq 0$ and $U(\lambda_{k+1}) \geq M_l$ **then**
 $\{\tilde{\mu}_{x,q}^{n,l}, \tilde{y}_{x,q}^{n,l}\} = \{\tilde{\mu}_{x,q}^{n,*}(\lambda_{k+1}), \tilde{y}_{x,q}^{n,*}(\lambda_{k+1})\}$
 $M_l = U(\lambda_{k+1})$
 end if
 $k = k + 1$
 end while
 $\theta = \frac{M - M_l}{M_u - M_l}$
 $\{\tilde{\mu}_{x,q}^{n,*}, \tilde{y}_{x,q}^{n,*}\} = \theta\{\tilde{\mu}_{x,q}^{n,u}, \tilde{y}_{x,q}^{n,u}\} + (1 - \theta)\{\tilde{\mu}_{x,q}^{n,l}, \tilde{y}_{x,q}^{n,l}\}$
end if

5.2. Truncation for the Hard Bandwidth Constraint

Finally, a bandwidth-truncated policy $\hat{\pi}_X$ is derived from the optimal stationary policy π_X^* to satisfy the hard bandwidth constraint in Equation (6b). Before introducing the truncated policy, first denote $S(t)$ to be the set of sensors to be scheduled in slot t , and $|S(t)|$ is the number of sensors to be scheduled in slot t . Then, the construction of $\hat{\pi}_X$ is carried out as follows.

- In slot t , compute the scheduling set $S(t)$ according to the optimal stationary policy π_X^* .
- If $|S(t)| \leq M$, then $\hat{\pi}_X$ schedules all these sensors as π_X^* does.
- If $|S(t)| > M$, the hard bandwidth constraint is never satisfied. Therefore, $\hat{\pi}_X$ randomly chooses M out of $|S(t)|$ sensors to be scheduled in the current slot.

The following theorem guarantees the asymptotic performance of the truncated policy $\hat{\pi}_X$ compared with π_X^* on certain conditions.

Theorem 4. Suppose the age penalty function is concave, and let $\kappa = \frac{M}{N}$ be a constant. If all the sensors and their channels are identical, i.e., the power constraint and the channel transition

matrix are the same, then the truncated policy $\hat{\pi}_X$ and the optimal randomized policy π_X^* have the following property,

$$\lim_{N \rightarrow \infty} J(\hat{\pi}_X) - J(\pi_X^*) = 0.$$

The whole proof is provided in Appendix F.

6. Simulation Results

In this section, we provide simulation results to verify the performance of the proposed policy. First, we study the average age penalty performance with different types of sensors with different bandwidth constraint and AoI truncation threshold X . Next, we study the detailed scheduling decision of each sensor. The average performance is obtained by simulating 10^5 consecutive slots.

6.1. Average Age Penalty Performance

In this part, we demonstrate the average performance of our proposed policy. We consider 4-state channel system, i.e., $Q = 4$. The age penalty function is chosen as $f(x) = \ln(x)$ unless otherwise specified. The transition matrix P_n for each sensor is the same:

$$P_n = \begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.25 & 0.3 & 0.25 & 0.2 \\ 0.2 & 0.25 & 0.3 & 0.25 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{pmatrix}. \tag{21}$$

Denote $\{\eta_q\}$ to be the steady distribution of the channel state. We consider that for each channel state q , the energy consumption $w(q) = q$. According to [12], the optimal policy to minimize the average AoI performance when all the sensors are identical is a greedy policy, which schedules the M sensors with the largest AoI and consumes the average power for each sensor $\mathcal{E}_G = \frac{M}{N} \sum_{q=1}^Q \eta_q w(q)$. Therefore, define $\rho_n = \frac{\mathcal{E}_n}{\mathcal{E}_G}$ to be the power constraint factor which describes the effects of power consumption constraint \mathcal{E}_n .

Figure 2 demonstrates the average age penalty performance of the proposed policy $\hat{\pi}_X$ as a number of sensors N , with bandwidth constraint $M = \{5, 15\}$, compared with the lower bound, the relaxed optimal policy π_X^* and the greedy policy. Set the threshold $X = \lceil \frac{20N}{M} \rceil$, where $\lceil \cdot \rceil$ is ceiling function. We assume that the probability of packet loss for each sensor is the same, denoted by ε :

$$\varepsilon = [0.1, 0.3, 0.2, 0.4].$$

The power constraint factor of sensor n is $\rho_n = 0.2 + \frac{1.4(n-1)}{N-1}$.

As seen in Figure 2, the proposed truncated policy performs closely to the relaxed optimal policy and the lower bound, and outperforms the greedy policy especially when N is large. According to Figure 2, the age penalty decreases by 18% and 23% from the greedy policy with $N = 60$ sensors when $M = 5$ and $M = 15$, respectively, under proposed policy. Moreover, as the threshold X becomes large, the difference between the average performance following policy π_X^* and the lower bound becomes indistinguishable. Therefore, the asymptotic performance described in Theorem 3 can be verified.

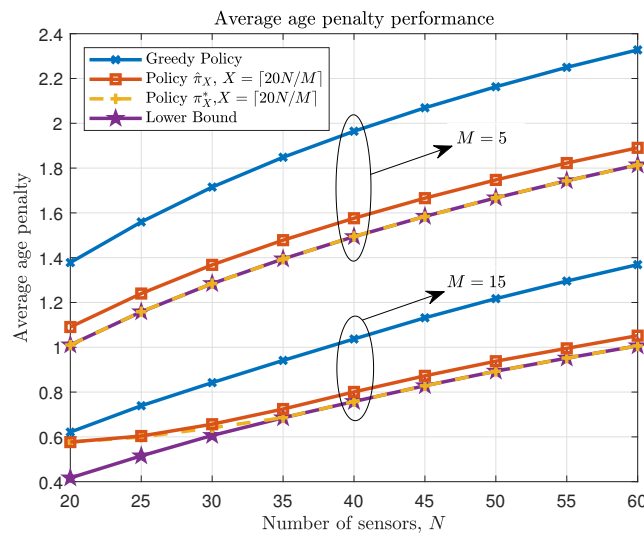


Figure 2. Average age penalty performance as a number of sensors N , $M = \{5, 15\}$.

Figure 3 compares the average performance of the proposed policy $\hat{\pi}_X$ with the AoI-minimum policy to verify the improvement of considering different penalty function. The AoI-minimum policy is similar to the one proposed in our previous work [22] with the consideration of packet loss. The bandwidth constraint $M = 2$. The packet loss probability $\varepsilon = [0.25, 0.35, 0.65, 0.8]$, and the energy consumption $w(q) = 2^q$. Here the age penalty function is chosen as $f(x) = x^2$. From Figure 3, we can see that the AoI-minimum policy cannot guarantee a good age penalty performance. Thus, it is necessary to consider different demand for data freshness to achieve better performance.

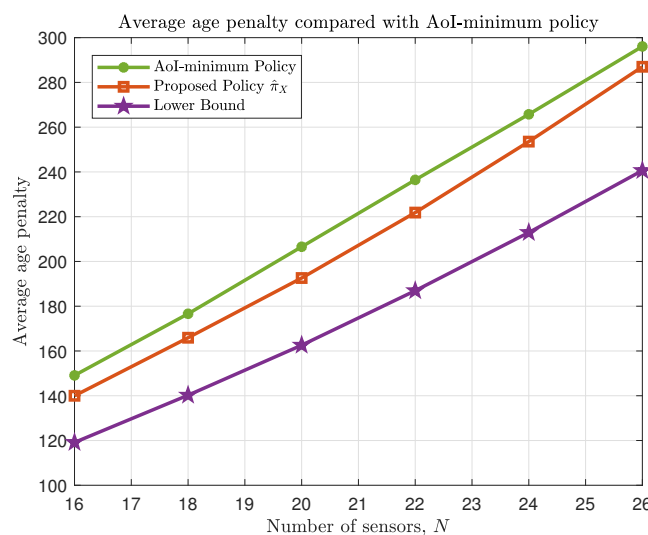


Figure 3. Average age penalty performance as a number of sensors N compared with AoI-minimum policy.

Figures 4 and 5 verify the asymptotic performance of the proposed policy with different age penalty function and different packet loss probability respectively when M/N is a constant in symmetric networks. From both figures, it can be seen that the difference of average age penalty under proposed policy and the lower bound becomes small as N increases. Thus, the asymptotic performance described in Theorem 4 can be verified.

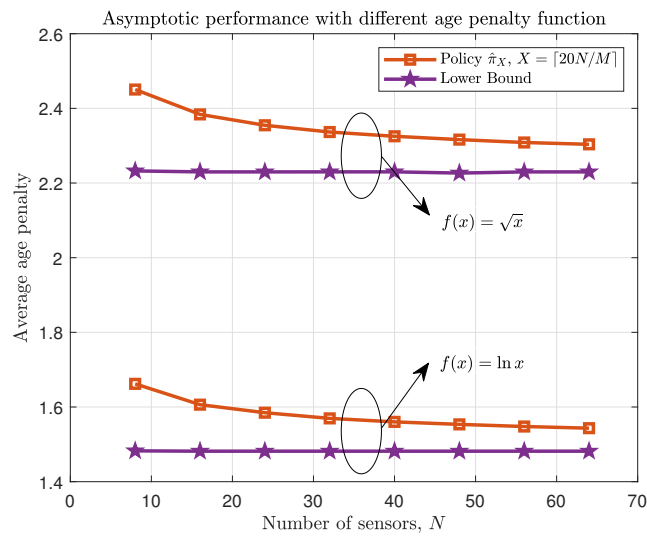


Figure 4. Asymptotic average age penalty performance with different age penalty function $f(x) = \{\ln x, \sqrt{x}\}$. $\kappa = \frac{1}{8}$.

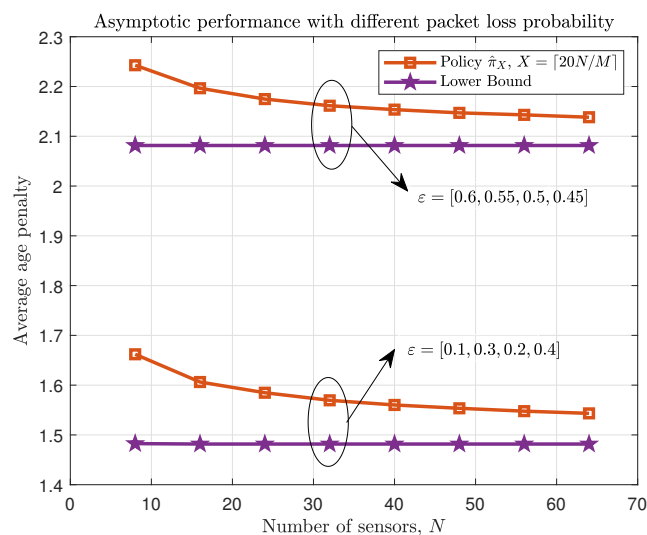


Figure 5. Asymptotic average age penalty performance with different packet loss probability. $\epsilon = \{[0.1, 0.3, 0.2, 0.4], [0.6, 0.55, 0.5, 0.45]\}$. $\kappa = \frac{1}{8}$.

6.2. Sensor Level Analysis and Threshold Structure

Next, we analyze the scheduling decision of each sensor and their corresponding age penalty to provide some insights of optimal scheduling policies. We consider a system with $N = 8$ and $M = 2$. The transition matrix of channel state is the same as Equation (21), and power consumption $w(q) = q$. We set the threshold $X = 80$ to compute the proposed policy.

First we consider the system with $Q = 4$ and age penalty function $f(x) = \ln x$. Figure 6 analyzes how the power constraint influences age penalty of each sensor. The power constraint of sensor n is $\rho_n = 0.2n$. From Figure 6, we can see that the proposed policy outperforms the greedy policy when the required power consumption is scarce, and performs similarly or a little worse when the factor $\rho_n > 1$. This implies that our proposed policy chooses a more proper power allocation based on current channel state and AoI than the greedy policy by stimulating sources with scarce power budgets to be scheduled in better channel states.

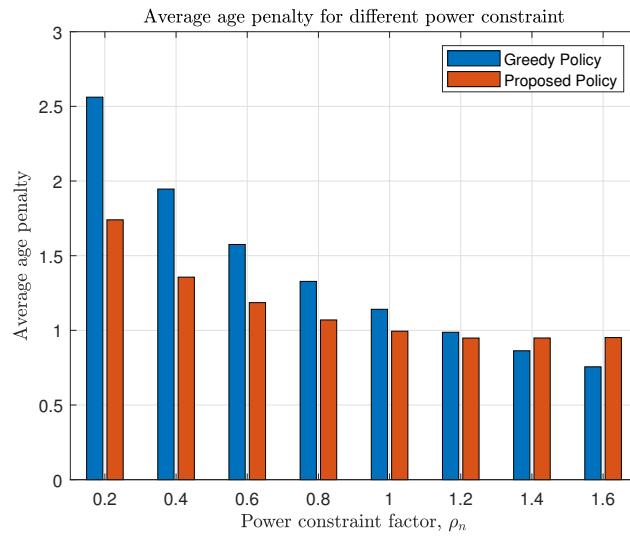


Figure 6. Each sensor’s average penalty with different power constraint factor $\rho_n = 0.2n$.

As the packet loss influences age penalty as well, Figure 7 considers sensors with different probability of packet loss, which can be written out as the following matrix $\{\varepsilon_{n,q}\}$:

$$\varepsilon = \begin{pmatrix} 0.05 & 0.1 & 0.15 & 0.2 \\ 0.1 & 0.15 & 0.2 & 0.25 \\ 0.15 & 0.2 & 0.25 & 0.3 \\ 0.2 & 0.25 & 0.3 & 0.35 \\ 0.25 & 0.3 & 0.35 & 0.4 \\ 0.3 & 0.35 & 0.4 & 0.45 \\ 0.35 & 0.4 & 0.45 & 0.5 \\ 0.4 & 0.45 & 0.5 & 0.55 \end{pmatrix}. \tag{22}$$

We fix the power constraint factor $\rho_n = 0.6$ for all sensors. Figure 7 shows that the average age penalty increases with the probability of packet loss. Moreover, the proposed policy combats with the packet loss better than the greedy policy as the proposed policy considers $\varepsilon_{n,q}$ when solving the LP problem, but the greedy policy does not.

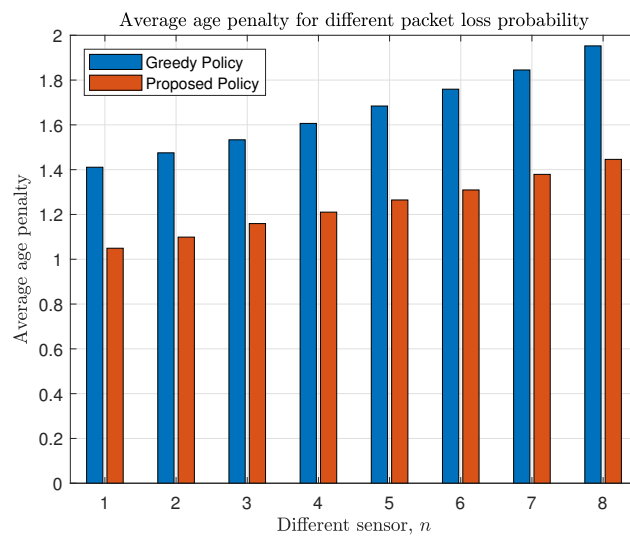


Figure 7. Each sensor’s average penalty with different packet loss probability $\varepsilon_{n,q} = 0.05(n + q - 1)$.

Next, we verify the threshold structure of the optimal scheduling policy. Figure 8 demonstrates the effect of bandwidth and packet loss on the scheduling threshold. The power constraint factor $\rho_n = 0.6, \forall n$, and the packet loss probability is the same as Equation (22). Subfigures (a–c) demonstrate three of these sensors whose packet loss probability is as the title. For each of the three sensors, subfigures (d–f) consider the single-sensor system without bandwidth constraint and display their scheduling probability, respectively. Moreover, Figure 8 lists some of the thresholds given channel state q , e.g., in subfigure (a), the threshold of channel state $q = 3$ is $x = 7$, and the corresponding optimal scheduling probability is $\xi_{7,3} = 0.9963$. From Figure 8, first we can see that all the six figures verify the non-decreasing property of the scheduling probability with AoI $x(t)$ described in Lemma 2. Second, subfigures (a–c) demonstrate that the sensor with higher packet loss probability also has higher scheduling threshold. This implies that the sensors with more reliable channel should be given higher priority to scheduling than unreliable ones to minimize the average age penalty, since scheduling the more reliable channel under the same AoI is more likely to reduce the current age penalty. Third, by comparing subfigures (a) and (d), (b) and (e), and (c) and (f), the scheduling threshold varies more significantly for different channel states if there exists no bandwidth constraint. The sensors tend to update more often when the channel state is good, and idle when the channel state is bad. This is because the sensors can choose to update packets more frequently in good channel state to both save energy and increase the success probability of transmission without bandwidth constraint.

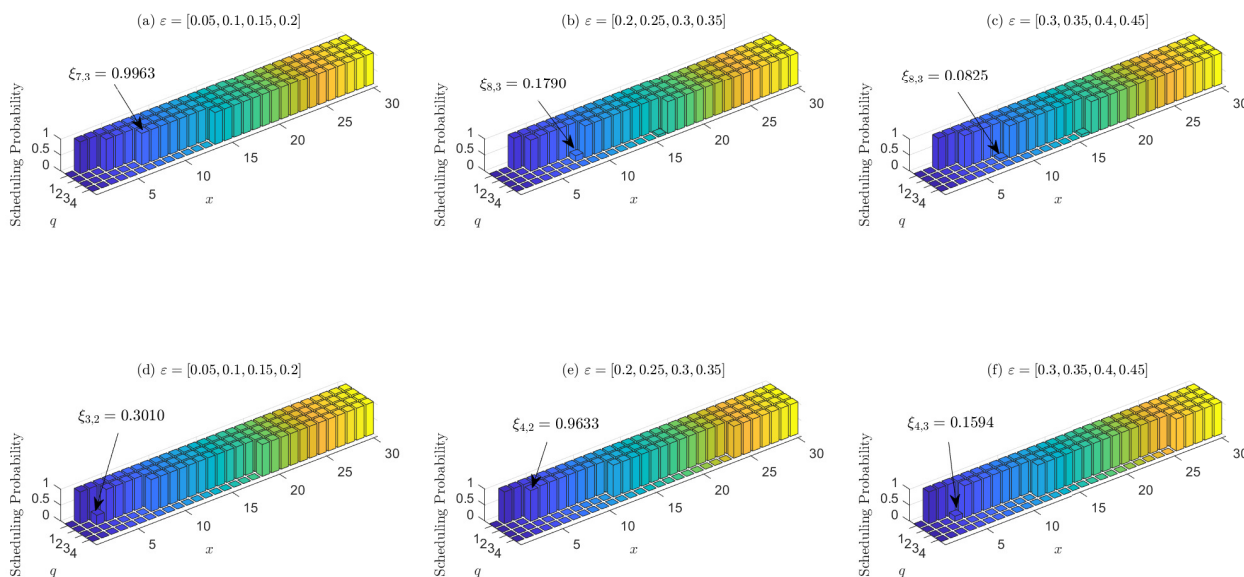


Figure 8. Subfigures (a–c) demonstrate the scheduling probability of sensors whose packet loss probability is as the title when $N = 8$ and $M = 2$. Subfigures (d–f) demonstrate their scheduling probability in single-sensor system respectively. All the power constraint is $\rho_n = 0.6$.

Finally, we study the effects of age penalty function on threshold structure. Here, we consider a system with $Q = 2, \rho_n = 0.2n$ and three different penalty function, i.e., $f(x) = \ln x, f(x) = x$, and $f(x) = x^2$ in Figure 9. We plot the scheduling decision of the sixth sensor. As is depicted in Figure 9, as the system has a higher restriction on data freshness such as exponential or quadratic function, the difference between thresholds of different states becomes small. In such situations, channel states play a weaker role because waiting for another slot to schedule tends to have unbearable age penalty.

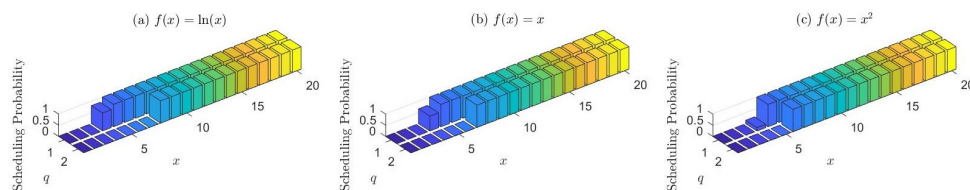


Figure 9. Threshold comparison with different age penalty function.

7. Conclusions

In this paper, we consider the multi-sensor scheduling problem through an error-prone Markovian channel state. Through relaxing and decoupling, we propose a truncated policy to satisfy both the bandwidth and power constraints to minimize the average age penalty of all sensors in infinite horizon. We prove the asymptotic performance of the truncated policy in symmetric networks when the age penalty function is concave by choosing a sufficiently large threshold X . Through theoretical analysis and numerical simulations, we find that the age penalty function, packet loss probability, bandwidth constraint, and power constraint work altogether to influence the optimal scheduling decisions. Those who have more reliable channel state and enough power consumption tend to have higher scheduling priority.

Author Contributions: Conceptualization, Y.C. and H.T.; methodology, Y.C.; validation, Y.C. and H.T.; formal analysis, Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, H.T., J.W., and J.S.; supervision, J.W. and J.S.; funding acquisition, J.W. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key R&D Program of China under Grant 2017YFE0112300, Beijing National Research Center for Information Science and Technology under Grant BNR2019RC01014 and BNR2019TD01001.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

First, we notice that Problem 2 is equivalent to the following optimization problem, where we further introduce variables v_n to denote the local bandwidth constraint of sensor n .

Problem A1 (Equivalent Relaxed Primal Scheduling Problem).

$$Age_R^* = \min_{\pi \in \Pi_{CP, \{v_1, \dots, v_N\}}} \lim_{T \rightarrow \infty} \frac{1}{NT} \mathbb{E}_\pi \left[\sum_{n=1}^N \sum_{t=1}^T f(x_n(t)) \right], \tag{A1a}$$

$$s.t. \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=1}^T u_n(t) \right] \leq v_n, \forall n, \tag{A1b}$$

$$\sum_{n=1}^N v_n \leq M, \tag{A1c}$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \forall n, \tag{A1d}$$

$$0 \leq v_n \leq 1, \forall n. \tag{A1e}$$

For each feasible fixed local bandwidth constraint vector $\{v_n\}$, we can transfer the above problem into the following one by removing constraint Equations (A1c) and (A1e).

Problem A2 (Relaxed Problem for Fixed $\{v_n\}$).

$$Age_R^*(v_1, \dots, v_n) = \min_{\pi \in \Pi_{CP}} \lim_{T \rightarrow \infty} \frac{1}{NT} \mathbb{E}_\pi \left[\sum_{n=1}^N \sum_{t=1}^T f(x_n(t)) \right], \tag{A2a}$$

$$s.t. \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=1}^T u_n(t) \right] \leq v_n, \forall n, \tag{A2b}$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_\pi \left[\sum_{t=1}^T u_n(t) w(q_n(t)) \right] \leq \mathcal{E}_n, \forall n. \tag{A2c}$$

Then, according to the work in [25], the optimal policy of the Problem 6 given $\{v_n\}$ can be decoupled into several local policies. This is somewhat intuitive as the constraints and objective function of Problem 6 are decoupled for each sensor n . As for each feasible $\{v_n\}$, the optimal scheduling policy can be decoupled, recall that $Age_R^* = \min_{v_1, \dots, v_N} Age_R^*(v_1, \dots, v_N)$, when $\{v_n\}$ takes the optimum value $\{v_n^*\}$, the property also holds.

Appendix B. Proof of Lemma 1

Before we proceed to the proof, we make two definitions. For any two states $s = (x, q)$ and $s' = (x', q) \in \mathbb{S}$, define a partial order \leq_s . We state that $s \leq_s s'$ if and only if $x \leq x'$. Moreover, we also define a partial order \leq_a on the action space \mathbb{A} . We state that $u \leq_a u'$ if and only if $u \leq u'$.

The monotonicity of the optimal action on the state space is true if the four following conditions hold:

1. If $s \leq_s s'$, $c(s, u) \leq c(s', u)$ for any $u \in \mathbb{A}$;
2. If $s \leq_s s'$, for any $u \in \mathbb{A}$, $\sum_{s_+} \Pr(s_+|s, u) V(s_+) \leq \sum_{s_+} \Pr(s_+|s', u) V(s_+)$, where $V(s)$ is any monotone increasing function;
3. If $s \leq_s s'$ and $u \leq_a u'$, then $c(s, u) + c(s', u') \leq c(s', u) + c(s, u')$;
4. If $s \leq_s s'$ and $u \leq_a u'$, then:

$$\begin{aligned} & \sum_{s_+} \Pr(s_+|s, u) V(s_+) + \sum_{s_+} \Pr(s_+|s', u') V(s_+) \\ & \leq \sum_{s_+} \Pr(s_+|s', u) V(s_+) + \sum_{s_+} \Pr(s_+|s, u') V(s_+); \end{aligned}$$

where $s, s' \in \mathbb{S}$, $u, u' \in \mathbb{A}$. $s^+ \in \mathbb{S}$ is the next state, and $c(s, u)$ denotes the one-step cost given the state s and action u .

Next, we consider a discounted cost MDP over a finite horizon:

$$\min_{\pi \in \Pi_{CP}} \mathbb{E}_\pi \left[\sum_{k=1}^T \beta^k c(s(k), u(k)) \right], \tag{A3}$$

where β is discounted factor.

And the corresponding Bellman equation is

$$V_{T, \beta}(s) = \min_{u \in \mathbb{A}} c(s, u); \tag{A4}$$

$$V_{k, \beta}(s) = \min_{u \in \mathbb{A}} [c(s, u) + \beta \sum_{s_+} \Pr(s_+|s, u) V_{k+1, \beta}(s_+)], k = 1, 2, \dots, T - 1. \tag{A5}$$

If the above four conditions are satisfied and the corresponding Bellman function is monotone increasing, then the one-step cost $c(s, u) = c_x(x, q, u) + v c_E(x, q, u)$ is monotone and sub-modular in s and u , which shows there exists an optimal monotone policy for any

finite-time horizon MDP. Using the vanishing discount approach in Theorem 5.5.4 in [26], the property of monotonicity is propagated to the time-average MDP.

Before verifying the above four conditions, first we introduce the following lemma to ensure $V_{k,\beta}(x, q)$ is monotone increasing with x , whose proof is provided in Appendix C.

Lemma A1. For fixed channel state q and discounted factor β , $V_{k,\beta}(x, q)$ is monotone increasing with x .

Therefore, we only need to show the decoupled unconstrained problem satisfies the above four conditions.

Notice that the one-step cost can be computed by

$$c(s, u) = f(x) + \lambda u + v w(q)u. \tag{A6}$$

According to the definition of partial order \leq_s and \leq_a , and Equation (A6), we can easily verify condition 1 and 3.

Also, we have:

$$\sum_{s_+} \Pr(s_+ | s, u) V(s_+) = \begin{cases} \sum_{q'} p_{qq'} V(x + 1, q'), & u = 0; \\ \sum_{q'} [p_{qq'}(1 - \varepsilon_q) V(1, q') + p_{qq'} \varepsilon_q V(x + 1, q')], & u = 1. \end{cases} \tag{A7}$$

According to Equation (A7) and the fact that $V(x, q) < V(x', q), \forall x < x'$, it is also feasible to verify that both condition 2 and 4 hold.

Appendix C. Proof of Lemma A1

In this part, we will prove that $V_{k,\beta}(x, q)$ in the finite-time horizon MDP is an increasing function with x . The key method of the proof is through induction.

Notice that $V_{T,\beta}(x, q) = \min_{u \in \mathbb{A}} c(s, u) = \min_{u \in \mathbb{A}} f(x) + \lambda u + v w(q)u$ is increasing with x . Suppose that $V_{t,\beta}(x, q)$ is increasing with $x, \forall t \geq k$. For $\forall x_1 < x_2$, we have

$$c_x(x_1, q, u) < c_x(x_2, q, u), c_E(x_1, q, u) = c_E(x_2, q, u).$$

Denote $J_{k-1,\beta}(x, q, u)$ to be the expected discounted cost if take action u at state (x, q) . Then,

$$\begin{aligned} J_{k-1,\beta}(x_1, q, 0) &= c_x(x_1, q, 0) + \beta \sum_{q'} p_{qq'} V_{k,\beta}(x_1 + 1, q) \\ &< c_x(x_2, q, 0) + \beta \sum_{q'} p_{qq'} V_{k,\beta}(x_2 + 1, q) \\ &= J_{k-1,\beta}(x_2, q, 0). \end{aligned}$$

Similarly, we can derive that $J_{k-1,\beta}(x_1, q, 1) < J_{k-1,\beta}(x_2, q, 1)$. According to the Bellman equation Equations (A4) and (A5), the value function $V_{k-1,\beta}(x, q)$ can be computed as follows,

$$V_{k-1,\beta}(x, q) = \min_{u \in \mathbb{A}} J_{k-1,\beta}(x, q, u).$$

From the above, $J_{k-1,\beta}(x_1, q, u) < J_{k-1,\beta}(x_2, q, u), \forall u \in \mathbb{A}$. Thus, $V_{k-1,\beta}(x_1, q) < V_{k-1,\beta}(x_2, q)$. Therefore, through induction, we verify that $V_{k,\beta}(x, q)$ is increasing with x .

Appendix D. Derivation of Problem 4

First, make the following variable substitution,

$$\mu'_{x,q} = \begin{cases} \mu_{x,q}, & x < X; \\ \sum_{x'=X}^{\infty} \mu_{x',q}, & x = X. \end{cases} \tag{A8}$$

Similarly, let

$$y'_{x,q} = \begin{cases} y_{x,q}, & x < X; \\ \sum_{x'=X}^{\infty} y_{x',q}, & x = X. \end{cases} \quad (\text{A9})$$

Instead of solving $\mu_{x,q}$ and $y_{x,q}$ in Theorem 2, we solve the new finite variables $\mu'_{x,q}$ and $y'_{x,q}$. Therefore, the constraints Equation (13b), Equation (13c), Equation (13e), and Equation (13f) become Equation (16b), Equation (16c), Equation (16f), and Equation (16g), respectively. Equation (16d) corresponds to Equation (13d) when $x \leq X - 1$. Otherwise, sum all the Equation (13d) from X to infinity, i.e.,

$$\begin{aligned} \sum_{x=X}^{\infty} \mu_{x,q'} &= \sum_{x=X}^{\infty} \sum_{q=1}^Q p_{qq'} [\mu_{x-1,q} - (1 - \varepsilon_q) y_{x-1,q}] \\ &\stackrel{(a)}{=} \sum_{q=1}^Q p_{qq'} [\mu_{X-1,q} - (1 - \varepsilon_q) y_{X-1,q}] \\ &\quad + \sum_{x=X}^{\infty} \sum_{q=1}^Q p_{qq'} [\mu_{x,q} - (1 - \varepsilon_q) y_{x,q}] \\ &= \sum_{q=1}^Q p_{qq'} [\mu_{X-1,q} - (1 - \varepsilon_q) y_{X-1,q}] \\ &\quad + \sum_{q=1}^Q p_{qq'} \left[\sum_{x=X}^{\infty} \mu_{x,q} - (1 - \varepsilon_q) \sum_{x=X}^{\infty} y_{x,q} \right], \end{aligned}$$

where (a) holds because we separate $x = X$ from the sum.

Invoking Equations (A8) and (A9), the above equality is equivalent to

$$\begin{aligned} \mu'_{X,q'} &= \sum_{q=1}^Q p_{qq'} [\mu'_{X-1,q} - (1 - \varepsilon_q) y'_{X-1,q}] + \sum_{q=1}^Q p_{qq'} [\mu'_{X,q} - (1 - \varepsilon_q) y'_{X,q}] \\ &= \sum_{q=1}^Q p_{qq'} [\mu'_{X-1,q} + \mu'_{X,q} - (1 - \varepsilon_q) (y'_{X-1,q} + y'_{X,q})], \end{aligned}$$

which is exactly equal to Equation (16e).

Notice that the optimal action is to schedule when $x \geq X$, i.e., $\xi_{x,q}^* = 1$. Thus, $y_{x,q}^* = \mu_{x,q}^*, \forall x \geq X$. Therefore, we have

$$\mu'_{X,q} = y'_{X,q'} \quad (\text{A10})$$

which is equal to Equation (16h).

By now, through variable substitution, we have verified the constraints of the finite LP problem are equivalent to the ones of the original optimization problem. Therefore, the

optimal solution to the finite LP problem is also feasible to the original problem. Now for the objective function, we have

$$\begin{aligned}
 & \sum_{q=1}^Q \sum_{x=1}^{\infty} (f(x)\mu_{x,q} + \lambda y_{x,q}) \\
 &= \sum_{q=1}^Q \sum_{x=1}^{X-1} (f(x)\mu_{x,q} + \lambda y_{x,q}) + \sum_{q=1}^Q \sum_{x=X}^{\infty} (f(x)\mu_{x,q} + \lambda y_{x,q}) \\
 &\geq \sum_{q=1}^Q \sum_{x=1}^{X-1} (f(x)\mu_{x,q} + \lambda y_{x,q}) + \sum_{q=1}^Q \sum_{x=X}^{\infty} (f(X)\mu_{x,q} + \lambda y_{x,q}) \\
 &= \sum_{q=1}^Q \sum_{x=1}^{X-1} (f(x)\mu'_{x,q} + \lambda y'_{x,q}) + \sum_{q=1}^Q (f(X)\mu'_{X,q} + \lambda y'_{X,q}) \\
 &= \sum_{q=1}^Q \sum_{x=1}^X (f(x)\mu'_{x,q} + \lambda y'_{x,q}).
 \end{aligned}$$

Therefore, we have verified that the decoupled single-sensor problem is approximate to the LP problem, which has the same constraint but can be the lower bound of the original problem.

Appendix E. Proof of Theorem 3

Our aim is to bound $J_{\infty}(\pi^*) - J_X(\pi_X^*)$. Recall that π^* is the optimal solution to the primal problem. Thus, the difference can be bounded by $J_{\infty}(\pi_X^*) - J_X(\pi_X^*)$.

To further bound the above term, first denote $\rho_{x,q}^X$ and $\zeta_{x,q}^X$ to be the distribution and scheduling probability of state (x, q) under policy π_X^* , i.e.,

$$\rho_{x,q}^X = \mathbb{E}_{\pi_X^*} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{I}((x(t), q(t)) = (x, q)) \right], \tag{A11}$$

$$\zeta_{x,q}^X = \mathbb{E}_{\pi_X^*} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{I}(u(t) = 1 | (x(t), q(t)) = (x, q)) \right], \tag{A12}$$

where $\mathbb{I}(\cdot)$ is indicating function.

Let $\gamma_{x,q}^X = \rho_{x,q}^X \zeta_{x,q}^X$. Recalling the approximation solution to Problem 4, we have

$$\rho_{x,q}^X = \tilde{\mu}_{x,q}^*, \quad \forall x < X, \quad \sum_{x=X}^{\infty} \rho_{x,q}^X = \tilde{\mu}_{X,q}^*; \tag{A13}$$

$$\gamma_{x,q}^X = \tilde{y}_{x,q}^*, \quad \forall x < X, \quad \sum_{x=X}^{\infty} \gamma_{x,q}^X = \tilde{y}_{X,q}^*. \tag{A14}$$

Then, we can bound $J_{\infty}(\pi_X^*) - J_X(\pi_X^*)$ as

$$\begin{aligned}
 & J_\infty(\pi_X^*) - J_X(\pi_X^*) \\
 &= \mathbb{E}_{\pi_X^*} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (f(x(t)) + \lambda u(t)) \right] - J_X(\pi_X^*) \\
 &= \mathbb{E}_{\pi_X^*} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{q=1}^Q \sum_{x=1}^\infty f(x) \mathbb{I}((x(t), q(t)) = (x, q)) \right. \\
 &\quad \left. + \lambda \mathbb{I}((x(t), q(t)) = (x, q), u(t) = 1) \right] - J_X(\pi_X^*) \\
 &\stackrel{(a)}{=} \sum_{q=1}^Q \sum_{x=1}^\infty (f(x) \rho_{x,q}^X + \lambda \gamma_{x,q}^X) - J_X(\pi_X^*) \tag{A15} \\
 &= \sum_{q=1}^Q \sum_{x=1}^{X-1} (f(x) \rho_{x,q}^X + \lambda \gamma_{x,q}^X) + \sum_{q=1}^Q \sum_{x=X}^\infty (f(x) \rho_{x,q}^X + \lambda \gamma_{x,q}^X) \\
 &\quad - \sum_{q=1}^Q \sum_{x=1}^X (f(x) \tilde{\mu}_{x,q}^* + \lambda \tilde{\gamma}_{x,q}^*) \\
 &\stackrel{(b)}{=} \sum_{x=X}^\infty (f(x) - f(X)) \sum_{q=1}^Q \rho_{x,q}^X,
 \end{aligned}$$

where (a) holds because of Equations (A11) and (A12). (b) holds because of Equations (A13) and (A14).

Next, we upper bound term $\sum_{q=1}^Q \rho_{x,q}^X$. For simplicity, denote $\rho_x^X = [\rho_{x,1}^X, \rho_{x,2}^X, \dots, \rho_{x,Q}^X]^T$. Thus, we have

$$\rho_{x+1}^X = \alpha \rho_x^X, \quad \forall x \geq \tau_{\max}, \tag{A16}$$

where the matrix α can be computed according to Equation (14) and further bounded as follows,

$$\alpha = \begin{pmatrix} \varepsilon_1 p_{1,1} & \varepsilon_2 p_{2,1} & \dots & \varepsilon_Q p_{Q,1} \\ \varepsilon_1 p_{1,2} & \varepsilon_2 p_{2,2} & \dots & \varepsilon_Q p_{Q,2} \\ \dots & \dots & \dots & \dots \\ \varepsilon_1 p_{1,Q} & \varepsilon_2 p_{2,Q} & \dots & \varepsilon_Q p_{Q,Q} \end{pmatrix} \leq \varepsilon_{\max} P^T. \tag{A17}$$

Therefore, $\sum_{q=1}^Q \rho_{x,q}^X$ can be bounded as follows,

$$\begin{aligned}
 \sum_{q=1}^Q \rho_{x,q}^X &= \mathbf{1}^T \rho_x^X \\
 &\stackrel{(a)}{=} \mathbf{1}^T \alpha^{x-\tau_{\max}} \rho_{\tau_{\max}}^X \\
 &\stackrel{(b)}{\leq} \mathbf{1}^T \varepsilon_{\max}^{x-\tau_{\max}} (P^T)^{x-\tau_{\max}} \rho_{\tau_{\max}}^X \\
 &\stackrel{(c)}{=} \varepsilon_{\max}^{x-\tau_{\max}},
 \end{aligned} \tag{A18}$$

where $\mathbf{1}$ is all-one vector. (a) holds because of Equation (A16). (b) holds due to Equation (A17). (c) holds because $\mathbf{1}$ is the eigenvector of P .

Plugging Equation (A18) into Equation (A15), we have

$$\begin{aligned} J_\infty(\pi^*) - J_X(\pi_X^*) &\leq \sum_{x=X}^{\infty} k(\epsilon^{-x} - \epsilon^{-X})\epsilon_{\max}^{x-\tau_{\max}} \\ &= \sum_{x=X}^{\infty} k(\epsilon_{\max}^{\frac{x}{2}-\tau_{\max}} - \epsilon_{\max}^{x-\frac{X}{2}-\tau_{\max}}) \\ &= \frac{k\epsilon_{\max}^{\frac{X+1}{2}-\tau_{\max}}}{1 - \epsilon_{\max}}. \end{aligned}$$

Appendix F. Proof of Theorem 4

According to Lemma 2, the optimal policy for every decoupled single-sensor problem has the threshold structure. Let τ_q^n be the threshold of sensor n given channel state q . Denote $\Gamma_n = \max_q \tau_q^n - \min_q \tau_q^n$ to be the largest difference between different thresholds of sensor n , and $\Gamma = \max_n \Gamma_n$. As all the sensors are identical, Γ does not change with N . Moreover, let $\bar{S}(t) = \mathbb{E}_{\pi_X^*}[S(t)]$.

Suppose that the sensor n is not scheduled under $\hat{\pi}_X$ when $u_n(t) = 1$. Now consider the probability that it is still not scheduled in the next slot, which results from two reasons, i.e., it jumps into a state which has a higher scheduling threshold or there are still more than M sensors to be scheduled. Let p be the probability that the channel state jumps into a state having a higher scheduling threshold. Then, the probability of idling in the next slot, denoted by p_{idle} can be computed by

$$p_{\text{idle}} = p + (1 - p)\frac{N - M}{N} = p\frac{M}{N} + \frac{N - M}{N}.$$

Notice that p can be upper bounded as

$$p \leq \max_{n,q} \sum_{q'=1, q' \neq q}^Q p_{qq'}^n = \max_{n,q} (1 - p_{qq}^n).$$

Therefore, p_{idle} can be upper bounded by z , which can be computed as

$$p_{\text{idle}} \leq z = \frac{M}{N} \max_{n,q} (1 - p_{qq}^n) + \frac{N - M}{N}.$$

Therefore, it can be generalized that the probability that it is still not scheduled in the consecutive k slots is upper bounded by $z^{(k-\Gamma_n)^+}$, where $(\cdot)^+ = \max(\cdot, 0)$.

Now, we bound the different performance of $\hat{\pi}_X$ and π_X^* by introducing another policy $\tilde{\pi}_X$. Under $\tilde{\pi}_X$, when $|S(t)| > M$, all these sensors are scheduled like π_X^* , but add an extra penalty $a_x^n(t)$ for those sensors, which can be computed as

$$\begin{aligned} a_x^n(t) &= \sum_{k=0}^{\infty} z^{(k-\Gamma_n)^+} (f(x_n(t+k)) - f(k)) \\ &\stackrel{(a)}{\leq} \sum_{k=0}^{\infty} z^{(k-\Gamma_n)^+} (f(x_n(t)) - f(0)) \\ &\leq \left(\Gamma_n + \frac{1}{1-z} \right) f(x_n(t)) \\ &\leq \left(\Gamma + \frac{1}{1-z} \right) f(x_n(t)), \end{aligned}$$

where (a) holds because $f(x)$ is concave.

For simplicity, denote $A = \Gamma + \frac{1}{1-z}$, which is a constant once the channel transition matrix and power constraint are fixed.

As the age penalty function $f(x)$ is concave, the average age penalty cost under $\tilde{\pi}_X$ does not decrease compared with $\hat{\pi}_X$. Then, the difference between $J(\hat{\pi}_X)$ and $J(\pi_X^*)$ can be bounded as

$$\begin{aligned} J(\hat{\pi}_X) - J(\pi_X^*) &\leq J(\tilde{\pi}_X) - J(\pi_X^*) \\ &= \lim_{T \rightarrow \infty} \mathbb{E}_{\pi_X^*} \left[\frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T a_x^n(t) \frac{(|S(t)| - M)^+}{M} \right] \\ &\leq \lim_{T \rightarrow \infty} \mathbb{E}_{\pi_X^*} \left[\frac{1}{MNT} \sum_{n=1}^N \sum_{t=1}^T Af(x_n(t)) ||S(t)| - M| \right], \end{aligned}$$

Notice that when $x > X$, the optimal action is to schedule. Hence, the probability of $x_n(t) > X$ is upper bounded by $(\max_{n,q} \epsilon_{n,q})^{x-X}$. For simplicity, let $\rho = \max_{n,q} \epsilon_{n,q}$. Then, $\forall \epsilon > 0$, there exists $k = \lceil \frac{\ln \epsilon}{\ln \rho} \rceil$ such that the steady distribution $\mu_{x,q}^n, \forall x > X + k$ and n can be bounded as

$$\sum_{q=1}^Q \mu_{x,q}^n \leq \rho^{x-X} = \rho^k \rho^{x-(X+k)} \leq \epsilon \rho^{x-(X+k)}.$$

Then, we have

$$\begin{aligned} &J(\hat{\pi}_X) - J(\pi_X^*) \\ &\leq \lim_{T \rightarrow \infty} \frac{1}{MNT} \mathbb{E}_{\pi_X^*} \left[\sum_{n=1}^N \sum_{t=1}^T \mathbb{I}_{x_n(t) \leq X+k} Af(x_n(t)) ||S(t)| - M| \right] \\ &\quad + \frac{1}{MNT} \mathbb{E}_{\pi_X^*} \left[\sum_{n=1}^N \sum_{t=1}^T \mathbb{I}_{x_n(t) > X+k} Af(x_n(t)) ||S(t)| - M| \right] \\ &\leq \lim_{T \rightarrow \infty} \frac{Af(X+k)}{M} \mathbb{E}_{\pi_X^*} \left[\frac{1}{T} \sum_{t=1}^T (||S(t)| - \bar{S}(t)| + |\bar{S}(t) - M|) \right] \tag{A19} \\ &\quad + \frac{1}{MNT} \mathbb{E}_{\pi_X^*} \left[\sum_{n=1}^N \sum_{t=1}^T \mathbb{I}_{x_n(t) > X+k} ANf(x_n(t)) \right] \\ &= \lim_{T \rightarrow \infty} \frac{Af(X+k)}{\kappa N} \mathbb{E}_{\pi_X^*} \left[\frac{1}{T} \sum_{t=1}^T (||S(t)| - \bar{S}(t)| + |\bar{S}(t) - M|) \right] \\ &\quad + \frac{A}{\kappa} \mathbb{E}_{\pi_X^*} \left[\frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T \mathbb{I}_{x_n(t) > X+k} f(x_n(t)) \right]. \end{aligned}$$

As $f(x)$ is concave, it can be upper bounded by a linear function, i.e., $f(x) \leq mx, \forall x > X + k$. Therefore, the second term in the above inequality can be further bounded as

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{E}_{\pi_X^*} \left[\frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T \mathbb{I}_{x_n(t) > X+k} f(x_n(t)) \right] &= \frac{1}{N} \sum_{n=1}^N \sum_{x=X+k+1}^{\infty} \sum_{q=1}^Q \mu_{x,q}^n f(x) \\ &\leq \sum_{x=X+k+1}^{\infty} m \epsilon \rho^{x-(X+k)} x \\ &= m \epsilon \left[\frac{\rho}{1-\rho} (X+k+1) + \frac{\rho^2}{(1-\rho)^2} \right]. \end{aligned}$$

By choosing $\epsilon = N^{-1}$ and $k = \mathcal{O}(\ln N)$, the second term in Equation (A19) converges to 0 as N becomes large.

For the first term in Equation (A19), according to the work in [27], the expectation of $||S(t)| - \bar{S}|$ has the following property,

$$\lim_{T \rightarrow \infty} \mathbb{E}_{\pi_X^*} \left[\frac{1}{T} \sum_{t=1}^T ||S(t)| - \bar{S}(t)| \right] = \mathcal{O}(\sqrt{N}).$$

In addition, as policy π_X^* satisfies the relaxed bandwidth constraint, we have

$$\lim_{T \rightarrow \infty} \mathbb{E}_{\pi_X^*} \left[\frac{1}{T} \sum_{t=1}^T |\bar{S}(t) - M| \right] = 0.$$

Therefore, when $T \rightarrow \infty$, $J(\hat{\pi}_X) - J(\pi_X^*)$ can be upper bounded by

$$\mathcal{O}\left(\frac{X + \ln N}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{X + 1 + \ln N}{N}\right).$$

As the threshold X does not increase with N , $J(\hat{\pi}_X) - J(\pi_X^*)$ converges to 0 as N becomes infinite. Thus, the asymptotic performance of the truncated policy has been proven.

References

1. Kaul, S.; Gruteser, M.; Rai, V.; Kenney, J. Minimizing age of information in vehicular networks. In Proceedings of the 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, Salt Lake City, UT, USA, 27–30 June 2011; pp. 350–358.
2. Zhou, B.; Saad, W. Optimal Sampling and Updating for Minimizing Age of Information in the Internet of Things. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018; pp. 1–6.
3. Kaul, S.; Yates, R.; Gruteser, M. Real-time status: How often should one update? In Proceedings of the 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012; pp. 2731–2735.
4. Chen, K.; Huang, L. Age-of-information in the presence of error. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 2579–2583.
5. Costa, M.; Codreanu, M.; Ephremides, A. On the Age of Information in Status Update Systems with Packet Management. *IEEE Trans. Inf. Theory* **2016**, *62*, 1897–1910. [\[CrossRef\]](#)
6. Najm, E.; Yates, R.; Soljanin, E. Status updates through M/G/1/1 queues with HARQ. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 131–135.
7. Kosta, A.; Pappas, N.; Ephremides, A.; Angelakis, V. Age and value of information: Non-linear age case. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 326–330.
8. Wang, X.; Duan, L. Dynamic Pricing for Controlling Age of Information. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 962–966.
9. Sun, Y.; Uysal-Biyikoglu, E.; Yates, R.; Koksall, C.E.; Shroff, N.B. Update or wait: How to keep your data fresh. *IEEE Trans. Inf. Theory* **2017**, *63*, 7492–7508. [\[CrossRef\]](#)
10. Ceran, E.T.; Gündüz, D.; György, A. Average Age of Information With Hybrid ARQ Under a Resource Constraint. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 1900–1913. [\[CrossRef\]](#)
11. Bacinoglu, B.T.; Ceran, E.T.; Uysal-Biyikoglu, E. Age of information under energy replenishment constraints. In Proceedings of the 2015 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 1–6 February 2015; pp. 25–31.
12. Kadota, I.; Uysal-Biyikoglu, E.; Singh, R.; Modiano, E. Minimizing the Age of Information in broadcast wireless networks. In Proceedings of the 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 27–30 September 2016; pp. 844–851.
13. Maatouk, A.; Kriouile, S.; Assaad, M.; Ephremides, A. On the Optimality of the Whittle’s Index Policy for Minimizing the Age of Information. *arXiv* **2020**, arXiv:2001.03096.
14. Kadota, I.; Sinha, A.; Modiano, E. Scheduling Algorithms for Optimizing Age of Information in Wireless Networks with Throughput Constraints. *IEEE/ACM Trans. Netw.* **2019**, *27*, 1359–1372. [\[CrossRef\]](#)
15. Kadota, I.; Sinha, A.; Uysal-Biyikoglu, E.; Singh, R.; Modiano, E. Scheduling Policies for Minimizing Age of Information in Broadcast Wireless Networks. *IEEE/ACM Trans. Netw.* **2018**, *26*, 2637–2650. [\[CrossRef\]](#)
16. Bedewy, A.M.; Sun, Y.; Shroff, N.B. Age-optimal information updates in multihop networks. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 576–580.
17. Chen, X.; Gatsis, K.; Hassani, H.; Bidokhti, S.S. Age of Information in Random Access Channels. In Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21–26 June 2020.
18. Hsu, Y.; Modiano, E.; Duan, L. Age of information: Design and analysis of optimal scheduling algorithms. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 561–565.

19. Talak, R.; Kadota, I.; Karaman, S.; Modiano, E. Scheduling Policies for Age Minimization in Wireless Networks with Unknown Channel State. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 2564–2568.
20. Sun, Y.; Polyanskiy, Y.; Uysal, E. Sampling of the Wiener Process for Remote Estimation over a Channel with Random Delay. *IEEE Trans. Inf. Theory* **2019**, *66*, 1118–1135. [[CrossRef](#)]
21. Wang, J.; Ren, X.; Mo, Y.; Shi, L. Whittle Index Policy for Dynamic Multichannel Allocation in Remote State Estimation. *IEEE Trans. Autom. Control* **2020**, *65*, 591–603. [[CrossRef](#)]
22. Tang, H.; Wang, J.; Song, L.; Song, J. Minimizing Age of Information With Power Constraints: Multi-User Opportunistic Scheduling in Multi-State Time-Varying Channels. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 854–868. [[CrossRef](#)]
23. Wu, S.; Han, D.; Cheng, P.; Shi, L. Optimal Scheduling of Multiple Sensors over Lossy and Bandwidth Limited Channels. *arXiv* **2018**, arXiv:1804.05618.
24. Beutler, F.J.; Ross, K.W. Optimal policies for controlled Markov chains with a constraint. *J. Math. Anal. Appl.* **1985**, *112*, 236–252. [[CrossRef](#)]
25. Tajan, R.; Ciblat, P. Information-theoretic multi-user power adaptation in retransmission schemes. In Proceedings of the 2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Edinburgh, UK, 3–6 July 2016; pp. 1–5.
26. Hernández-Lerma, O.; Lasserre, J.B. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 30.
27. Diaconis, P.; Zabell, S. Closed form summation for classical distributions: Variations on a theme of de Moivre. *Stat. Sci.* **1991**, *6*, pp. 284–302. [[CrossRef](#)]