Low-Power Status Updates via Sleep-Wake Scheduling

Ahmed M. Bedewy, Yin Sun, *Senior Member, IEEE*, Rahul Singh, and Ness B. Shroff, *Fellow, IEEE*

Abstract—We consider the problem of optimizing the freshness of status updates that are sent from a large number of low-power sources to a common access point. The source nodes utilize carrier sensing to reduce collisions and adopt an asynchronized sleep-wake scheduling strategy to achieve a target network lifetime (e.g., 10 years). We use *age of information* (AoI) to measure the freshness of status updates, and design sleep-wake parameters for minimizing the weighted-sum peak AoI of the sources, subject to per-source battery lifetime constraints. When the sensing time (i.e., the time duration of carrier sensing) is zero, this sleep-wake design problem can be solved by resorting to a two-layer nested convex optimization procedure; however, for positive sensing times, the problem is non-convex. We devise a low-complexity solution to solve this problem and prove that, for practical sensing times that are short, the solution is within a small gap from the optimum AoI performance. When the mean transmission time of status-update packets is unknown, we devise a reinforcement learning algorithm that adaptively performs the following two tasks in an "efficient way": a) it learns the unknown parameter, b) it also generates efficient controls that make channel access decisions. We analyze its performance by quantifying its "regret", i.e., the sub-optimality gap between its average performance and the average performance of a controller that knows the mean transmission time. Our numerical and NS-3 simulation results show that our solution can indeed elongate the batteries lifetime of information sources, while providing a competitive AoI performance.

Index Terms—Age of information; Data freshness; Low-power; Sleep-wake scheduler; Lifetime; certainty equivalent

This paper was presented in part at ACM MobiHoc 2020 [1].

This work has been supported in part by ONR grants N00014-17-1-2417 and N00014-15-1-2166, Army Research Office grants W911NF-14-1-0368 and MURI W911NF-12-1-0385, National Science Foundation grants CNS-1446582, CNS-1421576, CNS-1518829, and CCF-1813050, and a grant from the Defense Thrust Reduction Agency HDTRA1-14-1-0058.

A. M. Bedewy is with the Department of ECE, The Ohio State University, Columbus, OH 43210 USA (e-mail: bedewy.2@osu.edu).

Y. Sun is with the Department of ECE, Auburn University, Auburn, AL 36849 USA (e-mail: yzs0078@auburn.edu).

R. Singh is with the Department of ECE, Indian Institute of Science, Bangalore 560012, India (e-mail: rahulsingh@iisc.ac.in).

N. B. Shroff is with the Department of ECE and the Department of CSE, The Ohio State University, Columbus, OH 43210 USA (e-mail: shroff.11@osu.edu).

1 Introduction

In applications such as networked monitoring and control systems, wireless sensor networks, autonomous vehicles, it is crucial for the destination node to receive timely status updates so that it can make accurate decisions. *Age of information* (AoI) [2] has been used to measure the freshness of status updates. More specifically, AoI is the age of the freshest update at the destination, i.e., it is the time elapsed since the freshest received update was generated. It should be noted that optimizing traditional network performance metrics, such as throughput or delay, do not attain the goal of timely updating. For instance, it is well known that AoI could become very large when the offered load is high or low [2]. In other words, AoI captures the information lag at the destination, and is hence more apt for achieving the goal of timely updates. Thus, AoI has recently attracted a lot of interests (see [3], [4] and references therein).

In a variety of information update systems, energy consumption is also a critical concern. For example, wireless sensor networks are used for monitoring crucial natural and human-related activities, e.g. forest fires, earthquakes, tsunamis, etc. Since such applications often require the deployment of sensor nodes in remote or hard-to-reach areas, they need to be able to operate unattended for long durations. Likewise, in medical sensor networks, battery replacement/recharging involves a series of medical procedures, leading to disutility to patients. Hence, energy consumption must be constrained in order to support a long battery life of 10-15 years [5][1]. For networks serving such real-time ap-

1. The computations performed in [5] are based on the specifications of commercially used devices. For example, the used transceiver is 2.4 GHz chipset from Chipcon, the CC2420 [6], and the used microcontroller is the Motorola 8-bit microcontroller MC9508RE8 [7]. For more detail about the supply voltage and current consumption, please see the aforementioned references.

plications, prolonging battery-life is crucial. Existing works on multi-source networks, e.g., [8]–[11], [11]–[20], focused exclusively on minimizing the AoI and overlooked the need to reduce power consumption. This motivates us to derive scheduling algorithms that achieve a trade-off between the competing tasks of minimizing AoI and reducing the energy consumption in multi-source networks.

Additionally, some status-update systems consist of a large number (e.g., hundreds of thousands) of densely packed wireless nodes, which are serviced by a single access point (AP). Examples include massive machine-type communications [21]. The dataloads in such "dense networks" [21], [22] are created by applications such as home security and automation, oilfield and pipeline monitoring, smart agriculture, animal and livestock tracking, etc. This introduces high variability in the data packet sizes so that the transmission times of data packets are random. Thus, scheduling algorithms designed for time-slotted systems with a fixed transmission duration, are not applicable to these systems. Besides that, synchronized scheduler for time-slotted systems are feasible when there are relatively few sources and each source has sufficient energy. However, if there are a huge number of sources, the signaling overhead could be quite high. Since, each source may have limited energy and low traffic rate, the system could be highly inefficient. This motivates us to design asynchronized medium access protocols that coordinate the transmissions of multiple conflicting transmitters connected to a single AP.

Towards that end, we consider a wireless network with M sources that contend for channel access and communicate their update packets to an AP. Each source is equipped with a battery that may get charged by a renewable source of energy, e.g., solar. Moreover, each source employs a sleep-wake scheduling scheme [23] under which the source transmits a packet if the channel is idle; and sleeps if either: (i) The channel is busy, (ii) it has completed a packet transmission. This enables each source to save the precious battery energy by switching off when it is unlikely to gain channel access for packet transmissions. However, since a source cannot transmit during the sleep period, this causes the AoI to increase. We carefully design the sleep-wake parameters to minimize the weighted-sum peak age of the sources, while ensuring that the battery lifetime constraint of each source is satisfied.

1.1 Related Works

There have been significant recent efforts on analyzing the AoI performance of popular queueing service disciplines, e.g., the First-Come, First-Served (FCFS) [2] Last-Come, First-Served (LCFS) with and without preemption [24], and queueing systems with packet management [25]. In [18], [26]–[29], the age-optimality of Last-Generated, First-Served (LGFS)-type policies in multi-server and multi-hop networks was established, where it was shown that these policies can minimize any non-decreasing functional of the age processes. The design of data sampling and transmission in information update systems was investigated in [30], [31], where sampling policies were derived to minimize nonlinear age functions in single source systems. In [31], it was shown that a variety of information freshness metrics can be represented as monotonic functions of the age. The studies in [30], [31] were later extended to a multi-source scenario in [32], [33].

Designing scheduling policies for minimizing AoI in multi-source networks has recently received increasing attention, e.g., [8]–[17]. Of particular interest are those pertaining to designing distributed scheduling policies [8]–[13]. The work in [8] considered a slotted ALOHA-like random access scheme in which each node accesses the channel with a certain access probability. These probabilities were then optimized in order to minimize the AoI. However, the model of [8] allows multiple interfering users to gain channel access simultaneously, and hence allows for the collision. The authors in [9] generalized the work in [8] to a wireless network in which the interference is described by a general interference model. The Round Robin or Maximum Age First policy was shown to be (near) age-optimal for different system models, e.g., in [10]–[13], [18].

Carrier sensing distributed medium access mechanisms, e.g., Carrier Sense Multiple Access (CSMA), have been widely adopted in many wireless networks; see [34], [35] for a recent survey. There has been an interest in designing CSMA-based scheduling schemes that optimize the AoI [36], [37]. In [36], the authors designed an idealized CSMA (similar to that in [38]) to minimize the AoI with an exponentially distributed packet transmission times. In [37], the authors designed a slotted Carrier Sense Multiple Access/Collision-Avoidance (CSMA/CA) (similar to that in [39]) to minimize the broadcast age of information, which is defined, from a sender's perspective, as the age of the freshest successfully broadcasted packet. Contrary to these works, the sleep-wake scheduling scheme proposed by us emphasizes on reducing the cumulative energy consumption in multi-source networks in addition to minimizing the total weighted AoI. Moreover, in our study, transmission times are not necessarily random variables with some commonly used parametric density [36], or deterministic [37], but can be any generally

distributed random variables with finite mean.

1.2 Key Contributions

Our key contributions are summarized as follows:

- In our model, sources utilize an asynchronized sleep-wake scheduling strategy to achieve an extended battery lifetime. We aim at designing the mean sleeping period of each source, which controls its channel access probability, in order to minimize the total weighted average peak age of the sources while simultaneously meeting per-source battery lifetime constraints. Although, the aforementioned optimization problem is non-convex, we devise a solution. In the regime for which the sensing time is negligible compared to the packet transmission time, the proposed solution is near-optimal (Theorem 1 and Theorem 3). Our near-optimality results hold for general distributions of the packet transmission times.
- We propose an algorithm that can be easily implemented in many practical control systems. In particular, our solution requires the knowledge of only two variables in its implementation. These two variables are functions of the network parameters. An implementation procedure to compute these two variables is provided.
- As the ratio between the sensing time and the packet transmission time reduces to zero, we show that the age performance of our proposed algorithm is as good as that of the optimal synchronized scheduler (e.g., for time-slotted systems).
- Finally, since our solution is a function of the mean transmission time of data packets, the network operator needs to know this quantity in order to implement the algorithm. The transmission times however depend upon the environmental conditions, which in turn are hard to predict before the system operation begins. To overcome this challenge, we develop a reinforcement learning (RL) [40]–[42] algorithm that maintains an estimate of the (unknown) mean transmission time, and then utilizes this estimate in order to derive a solution that is "seemingly optimal" for the true system. We show that the regret of the proposed RL algorithm scales as $\tilde{O}(\sqrt{H})$,[2] where H is the operating time horizon.

2 Model and Formulation

2. $\tilde{O}$ hides factors that are logarithmic in H.

2.1 Network Model and Sleep-wake Scheduling

Consider a wireless network composed of M source nodes, each observing a time-varying signal. The sources generate update packets of the observed signals and send the packets to an access point (AP) over a shared spectrum band. If multiple sources transmit packets simultaneously, a packet collision occurs and these packet transmissions fail.

The sources use a sleep-wake scheduling scheme to access the shared spectrum, where each source switches between a sleep mode and a transmission mode over time, according the following rules: Upon waking from the sleep mode, a source first performs carrier sensing to check whether the channel is occupied by another source, as illustrated in Figure 1. The time duration of carrier sensing is denoted as t_s, which is sufficiently long to ensure a high sensing accuracy. If the channel is sensed to be busy, the source enters the sleep mode directly; otherwise, the source generates an update packet and sends it over the channel. The source hereafter goes back to the sleep mode.

In the above sleep-wake scheduling scheme, if two sources start transmitting within a time duration of t_s, then their sensing periods are overlapping and they may not be able to detect the transmission of each other. In order to obtain a robust system design, we consider that they cannot detect each other's transmission and a collision occurs. Upon completing a packet transmission, sources switch to the reception mode and wait for an acknowledgement (ACK) that indicates the outcome of their transmissions (successful transmission or collision). They then go back to the sleep mode.

A *sleep-wake cycle*, or simply a *cycle*, is defined as the time period between the ends of two successive packet transmission or collision events. Each cycle consists of an idle period and a transmission/collision period[3]. As depicted in Figure 1, the packet transmissions in Cycles 1-2 are successful, but a collision occurs in Cycle 3 because Sources 1 and 2 wake up within a short duration t_s.

We use $T_j, j \in \{1, 2, \ldots\}$ to represent the time incurred by the j-th packet transmission or collision event, which includes transmission/collision time and feedback delays.

3. To make the sleep-wake scheduling problem solvable analytically, we make several approximations. For example, in 802.11b frame structure, there exists a Short Inter-frame Space (SIFS) between the packet transmission frame and the ACK frame (i.e., the CTS frame). If another source wakes up during the SIFS, then it may not detect the transmission/ACK frames, leading to unexpected collisions. In our analytical model, such collision events are omitted. In other words, we suppose that each cycle must start with an idle period, where all sources are in the sleep mode, followed by a transmission/collision period. NS-3 simulation results will be provided in Section 5.2 to show that these approximations have a negligible impact on the age performance of our solution.

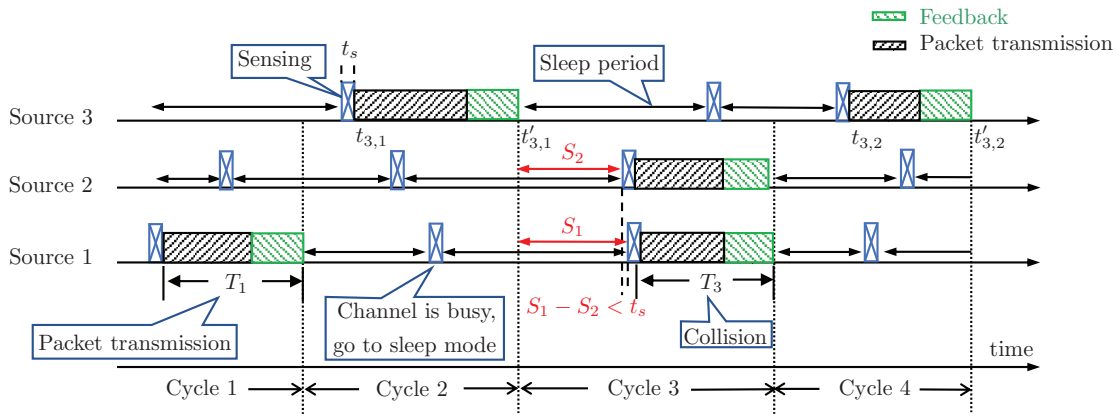

Figure 1: Illustration of the sleep-wake cycles. In Cycles 1-2, we have successful packet transmissions. Let S_1 and S_2 represent the remaining sleeping times of Sources 1 and 2, respectively, after a successful transmission. Then, a collision occurs in Cycle 3 because the difference between wake-up times of Sources 1 and 2 is less than t_s, i.e., $S_1 - S_2 < t_s$. As we can observe, each cycle consists of an idle period before a transmission/collision event.

For example, in Figure 1, T_1 is the time duration of the packet transmission event by Source 1, while T_3 is the time duration of the collision event between Source 1 and 2. We assume that the T_j's are i.i.d. for all transmission and collision events, with a general distribution. This assumption does not hold in practice. Nonetheless, NS-3 simulation results in Section 5.2 show that this assumption has a negligible impact on the performance of the proposed algorithm. When there is no confusion, we omit the subscript j of T_j for simplicity, and use T to denote the transmission/collision time, which is assumed to have a finite mean, i.e., $E[T] < \infty$. The sleep periods of source l are exponentially distributed random variables with mean value $\mathbb{E}[T]/r_l$ and are independent across sources and *i.i.d.* across time. Notice that, the sleep period parameter $r_l > 0$ has been normalized by the mean transmission time $\mathbb{E}[T]$. Let $\mathbf{r} = (r_1, \ldots, r_M)$ be the vector comprising of these sleep period parameters.

2.2 Total Weighted Average Peak Age

Let $U_l(t)$ represent the generation time of the most recently delivered packet from source l by time t. Then, the *age of information*, or simply the *age*, of source l is defined as [2]

$$\Delta_l(t) = t - U_l(t), \tag{1}$$

where $\Delta_l(t)$ is right-continuous. As shown in Figure 2, the age increases linearly with t, but is reset to a smaller value upon the delivery of a fresher packet. Observe that a small age $\Delta_l(t)$ indicates that the AP has a fresh status update packet that was generated at source l recently. Hence, it is desirable to keep $\Delta_l(t)$ small for all the sources.

Let us introduce some notations and definitions. Let i_l be the index of the i-th delivered packet from source l. We

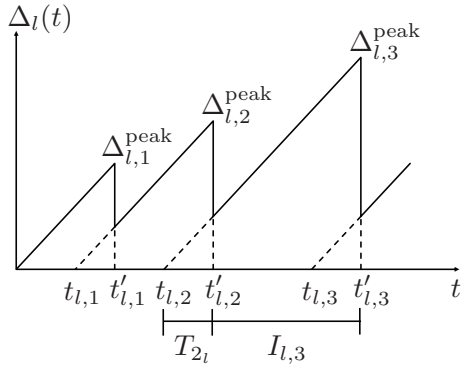

Figure 2: The age $\Delta_l(t)$ of source l.

use $t_{l,i}$ and $t'_{l,i}$ to denote the generation and delivery times, respectively, of the i-th delivered packet from source l, such that $t'_{l,i} - t_{l,i} = T_{i_l}$.[4] Let $I_{l,i} = t'_{l,i} - t'_{l,i-1}$ denote the i-th inter-departure time of source l, which satisfies $\mathbb{E}[I_{l,i}] = \mathbb{E}[I_l]$ for all i. The i-th peak age of source l, denoted by $\Delta^{\text{peak}}_{l,i}$, is defined as the AoI of source l right before the i-th packet delivery from source l. As shown in Figure 2, i.e., we have

$$\Delta^{\text{peak}}_{l,i} = \Delta_l(t'^{-}_{l,i}), \tag{2}$$

where $t'^{-}_{l,i}$ is the time instant just before the delivery time $t'_{l,i}$. One can observe from Figure 2 that the peak age is [25]

$$\Delta^{\text{peak}}_{l,i} = T_{(i-1)_l} + I_{l,i}. \tag{3}$$

Hence, the average peak age of source l is given by

$$\mathbb{E}[\Delta^{\text{peak}}_l] = \mathbb{E}[T] + \mathbb{E}[I_l], \tag{4}$$

where we omit the subscripts i and i_l as $I_{l,i}$'s and T_{i_l}'s are i.i.d. across time. The average peak age metric provides in-

4. A packet of a particular source is deemed delivered when the source receives the feedback.

formation regarding the worst case age, with the advantage of having a simpler formulation than the average age metric [25]. Thus, it is suitable for applications that have an upper bound restriction on AoI.

We now derive an expression for $\mathbb{E}[I_l]$. Let α_l be the probability of the event that the source l obtains channel access and successfully transmits a packet within a sleep-wake cycle. As shown in [23], one can utilize the memoryless property of exponential distributed sleep periods to get

$$\alpha_l = \frac{r_l e^{r_l \frac{t_s}{\mathbb{E}[T]}}}{e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \sum_{i=1}^{M} r_i}. \tag{5}$$

To keep the paper self-contained, we provide the derivation of (5) in Appendix A. Let N_l denote the total number of sleep-wake cycles between two subsequent successful transmissions of source l. Because the probability that source l obtains channel access and transmits successfully in a given cycle is α_l, N_l is geometrically distributed with mean $\frac{1}{\alpha_l}$. By this and (5), we get

$$\mathbb{E}[N_l] = \frac{e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \sum_{i=1}^{M} r_i}{r_l e^{r_l \frac{t_s}{\mathbb{E}[T]}}}. \tag{6}$$

An inter-departure time duration of source l is composed of N_l consecutive sleep-wake cycles. With a slight abuse of notation, let $\mathbf{cycle}_{l,k}$ denote the duration of the k-th sleep-wake cycle after a successful transmission of source l. Hence,

$$\mathbb{E}[I_l] = \mathbb{E}\left[\sum_{k=1}^{N_l} \mathbf{cycle}_{l,k}\right]. \tag{7}$$

Note that $\mathbf{cycle}_{l,k}$'s are i.i.d. across time. Moreover, since the event $(N_l = n)$ depends only on the history, N_l is a stopping time [43]. Hence, it follows from Wald's identity [44] that

$$\mathbb{E}[I_l] = \mathbb{E}[N_l]\mathbb{E}[\mathbf{cycle}], \tag{8}$$

where $\mathbb{E}[\mathbf{cycle}]$ is the mean duration of a sleep-wake cycle. Each cycle consists of an idle period and a transmission/collision time, see Figure 1. Using the memoryless property of exponential distribution, we observe that the idle period is the minimum of i.i.d. exponential random variables. Thus, it can be shown that the idle period in each cycle is exponentially distributed with mean value equal to $\mathbb{E}[T]/\sum_{i=1}^{M} r_i$, where $\mathbb{E}[T]/r_l$ is the mean of sleep periods of source l. Hence, we have

$$\mathbb{E}[\mathbf{cycle}] = \frac{\mathbb{E}[T]}{\sum_{i=1}^{M} r_i} + \mathbb{E}[T]. \tag{9}$$

Substituting the expressions for $\mathbb{E}[N_l]$ and $\mathbb{E}[\mathbf{cycle}]$ from (6) and (9), respectively, into (8), and (4), we obtain

$$\mathbb{E}[\Delta_l^{\text{peak}}] = \frac{e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \mathbb{E}[T]}{r_l} e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \left(1 + \sum_{i=1}^{M} r_i\right) + \mathbb{E}[T]. \tag{10}$$

In this paper, we aim to minimize the total weighted average peak age, which is given by

$$\sum_{l=1}^{M} w_l \mathbb{E}[\Delta_l^{\text{peak}}] = \sum_{l=1}^{M} \frac{w_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \mathbb{E}[T]}{r_l} e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \left(1 + \sum_{i=1}^{M} r_i\right) + \sum_{l=1}^{M} w_l \mathbb{E}[T], \tag{11}$$

where $w_l > 0$ is the weight of source l. These weights enable us to prioritize the sources according to their relative importance [9], [15].

2.3 Energy Constraint

Each source is equipped with a battery that can possibly be recharged by a renewable energy source, such as solar. In typical wireless sensor networks, sources have a much smaller power consumption in the sleep mode than in the transmission mode. For example, if the sensor is equipped with the radio unit TR 1000 from RF Monolithic [45], [46], the power consumption in the sleep mode is 15 μW while the power consumption in the transmission mode is 24.75 mW. Motivated by this, we assume that the energy dissipation during the sleep mode is negligible as compared to the power consumption in the transmission mode. Moreover, we assume that the sensing time duration t_s is much shorter than the transmission time and hence neglect the energy consumed during channel sensing. In Section 5.2, we show that these assumptions have a negligible effect on the performance of the proposed sleep-wake scheduling algorithm. Under these assumptions, the amount of energy used by a source is equal to the amount of energy consumed in packet transmissions and feedback receptions.

The energy constraint on source l is described by the following parameters: a) Initial battery level B_l, which denotes the initial amount of energy stored in the battery, b) Target lifetime D_l, which is the minimum time-duration that the source l should be active before its battery is depleted, c) Average energy replenishment rate[5] R_l, which is the rate at which the battery of source l receives energy from its energy

5. It is assumed that R_l is either known, or it can be estimated accurately.

source. If source l does not have access to an energy source, then we have $R_l = 0$. Define $P_{\text{max},l}$ for source l as

$$P_{\text{max},l} = \frac{B_l}{D_l} + R_l, \ \forall l, \tag{12}$$

where $P_{\text{max},l}$ is the maximum allowable power consumption of source l such that the target lifetime D_l is met.

For the sleep-wake scheduling mechanism under consideration, it has been shown in [23] that the fraction of time in which source l is in the transmission mode is given by

$$\sigma_l = \frac{[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^{M} r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{\sum_{i=1}^{M} r_i + 1}. \tag{13}$$

For the sake of completeness, the derivation of σ_l is provided in Appendix B. Let $P_{\text{avg},l}$ denote the average power consumption of source l in the transmission mode. Then the actual power consumption of source l, denoted by $P_{\text{act},l}$, is given by

$$P_{\text{act},l} = \sigma_l P_{\text{avg},l}, \ \forall l. \tag{14}$$

For source l to achieve its target lifetime D_l, we must have

$$P_{\text{act},l} \leq P_{\text{max},l}, \ \forall l. \tag{15}$$

Define $b_l \triangleq P_{\text{max},l}/P_{\text{avg},l}$ as the target power efficiency of source l. By using (13)-(14), the constraints in (15) can be rewritten as

$$\sigma_l = \frac{[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^{M} r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{\sum_{i=1}^{M} r_i + 1} \leq b_l, \ \forall l. \tag{16}$$

Because $\sigma_l \leq 1$, if $b_l \geq 1$, then constraint (16) is always satisfied.

2.4 Problem Formulation

Our goal is to find the optimal sleep-wake parameters $\mathbf{r}$ that minimizes the total weighted average peak age in (11), while simultaneously ensuring the energy constraints (16) for all sources. Dividing the objective function (11) by $\mathbb{E}[T]$, we obtain the following optimization problem: (Problem **1**)

$$\begin{aligned} \bar{\Delta}_{\text{opt}}^{\text{w-peak}} \triangleq \min_{r_l > 0} & \sum_{l=1}^{M} \frac{w_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{r_l} e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \left(1 + \sum_{i=1}^{M} r_i\right) + \\ & \sum_{l=1}^{M} w_l \\ \textbf{s.t.} \ & \frac{[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^{M} r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{\sum_{i=1}^{M} r_i + 1} \leq b_l, \forall l, \end{aligned} \tag{17}$$

where $\bar{\Delta}_{\text{opt}}^{\text{w-peak}}$ is the optimal objective value of Problem **1**. We will use $\bar{\Delta}^{\text{w-peak}}(\mathbf{r})$ to denote the objective value for given sleeping period parameters $\mathbf{r}$. One can notice from (17) that the optimal sleeping period parameters depend on the sensing time t_s and the mean transmission time $\mathbb{E}[T]$ only through their ratio $t_s/\mathbb{E}[T]$. This insight plays a crucial role in subsequent analysis of Problem **1**.

3 Optimal Sleep-wake Scheduling

When $t_s = 0$, although Problem **1** is non-convex, it can be solved by defining an auxiliary variable $y = \sum_{i=1}^{M} r_i + 1$ and applying a nested optimization algorithm: In the inner layer, we optimize r_l for a given y. Then, we write the optimized objective as a function of y. In the outer layer, we optimize y. It happens that the inner and outer layer optimization problems are both convex. The details can be found in Section 3.3.

However, this method does not work for positive sensing times $t_s > 0$ and Problem **1** becomes non-convex. Hence, it is challenging to optimize $\mathbf{r}$ for positive t_s. In this section, we develop a low-complexity closed-form solution which is shown to be near-optimal if the sensing time t_s is short as compared with the mean transmission time $\mathbb{E}[T]$. Our solution is developed by considering the following two regimes separately: (i) *Energy-adequate regime* denoted as $\sum_{i=1}^{M} b_i \geq 1$, where the condition $\sum_{i=1}^{M} b_i \geq 1$ means that the sources have a sufficient amount of total energy to ensure that at least one source is awake at any time, (ii) *Energy-scarce regime* represented by $\sum_{i=1}^{M} b_i < 1$, which indicates that the sources have to sleep for some time to meet the sources' energy constraints.

3.1 Energy-adequate Regime

In the energy-adequate regime $\sum_{i=1}^{M} b_i \geq 1$, our solution $\mathbf{r}^\star := (r_1^\star, \ldots, r_M^\star)$ is given as

$$r_l^\star = \min\{b_l, \beta^\star \sqrt{w_l}\} x^\star, \forall l, \tag{18}$$

where $x^\star$ and $\beta^\star$ are expressed in terms of the parameters $\{b_i, w_i\}_{i=1}^{M}, t_s/\mathbb{E}[T]$ as follows:

$$x^\star = \frac{-1}{2} + \sqrt{\frac{1}{4} + \frac{\mathbb{E}[T]}{t_s}}, \tag{19}$$

and $\beta^\star$ is the unique root of

$$\sum_{i=1}^{M} \min\{b_i, \beta^\star \sqrt{w_i}\} \ = 1. \tag{20}$$

The performance of the above solution $\mathbf{r}^\star$ is manifested in the following theorem:

Theorem 1 (**Near-optimality**). *If* $\sum_{i=1}^{M} b_i \geq 1$, *then the solution* $\mathbf{r}^\star$ (18) - (20) *is near-optimal for solving* (17) *when* $t_s/E[T]$ *is sufficiently small, in the following sense:*[6]

$$\left|\bar{\Delta}^{w\text{-}peak}(\mathbf{r}^\star) - \bar{\Delta}^{w\text{-}peak}_{opt}\right| \leq 2\sqrt{\frac{t_s}{\mathbb{E}[T]}}C_1 + o\left(\sqrt{\frac{t_s}{\mathbb{E}[T]}}\right), \quad (21)$$

where

$$C_1 = \sum_{i=1}^{M} \frac{w_i}{\min\{b_i, \beta^\star\sqrt{w_i}\}}. \quad (22)$$

Proof. See Appendix C.1. □

From Theorem 1, we can obtain the following corollary:

Corollary 2 (**Asymptotic optimality**). *If* $\sum_{i=1}^{M} b_i \geq 1$, *then the solution* $\mathbf{r}^\star$ (18) - (20) *is asymptotically optimal for Problem* ***1*** *in* (17) *as* $t_s/\mathbb{E}[T] \to 0$, *i.e.,*

$$\lim_{\frac{t_s}{\mathbb{E}[T]}\to 0}\left|\bar{\Delta}^{w\text{-}peak}(\mathbf{r}^\star) - \bar{\Delta}^{w\text{-}peak}_{opt}\right| = 0. \quad (23)$$

Moreover, the asymptotic optimal objective value of Problem ***1*** *as* $t_s/\mathbb{E}[T] \to 0$ *is*[7]

$$\lim_{\frac{t_s}{\mathbb{E}[T]}\to 0} \bar{\Delta}^{w\text{-}peak}_{opt} = \sum_{i=1}^{M}\left[\frac{w_i}{\min\{b_i, \beta^\star\sqrt{w_i}\}} + w_i\right]. \quad (24)$$

Proof. See Appendix C.1. □

3.2 Energy-scarce Regime

Now, we present a solution to Problem **1** in the energy-scarce regime $\sum_{i=1}^{M} b_i < 1$, and show it is near-optimal. The solution $\mathbf{r}^\star$ of the energy-scarce regime is again given by (18), where $x^\star$ and $\beta^\star$ are

$$x^\star = \frac{\min_l c_l}{1 - \sum_{i=1}^{M} b_i}, \ \beta^\star = \sum_{i=1}^{M} \frac{1}{\sqrt{w_i}}, \quad (25)$$

and

$$c_l = \frac{2b_l\left(1 - \sum_{i=1}^{M} b_i\right)^2}{Q_l}, \quad (26)$$

$$Q_l = b_l\left(1 - \sum_{i=1}^{M} b_i\right)^2 + \sqrt{b_l^2\left(1 - \sum_{i=1}^{M} b_i\right)^4 + 4b_l^2\left(1 - \sum_{i=1}^{M} b_i\right)^2\left(\sum_{i=1}^{M} b_i - b_l\right)\frac{t_s}{\mathbb{E}[T]}}. \quad (27)$$

6. We use the standard order notation: $f(h) = O(g(h))$ means $z_1 \leq \lim_{h\to 0} f(h)/g(h) \leq z_2$ for some constants $z_1 > 0$ and $z_2 > 0$, while $f(h) = o(g(h))$ means $\lim_{h\to 0} f(h)/g(h) = 0$.

7. Observe that, according to (24), the asymptotic optimal average peak age of source l is $(1/\min\{b_l, \beta^\star\sqrt{w_l}\} + 1)$ which decreases with the weight w_l. The weighted average peak age is $w_l(1/\min\{b_l, \beta^\star\sqrt{w_l}\} + 1)$ which increases with w_l. This phenomenon is reasonable and agrees with our expectation.

The near-optimality of the proposed solution (i.e., $\mathbf{r}^\star$) in the energy scarce regime is explained in the following theorem:

Theorem 3 (**Near-optimality**). *If* $\sum_{i=1}^{M} b_i < 1$, *then the solution* $\mathbf{r}^\star$ (18) *and* (25) - (27) *is near-optimal for solving* (17) *when* $t_s/\mathbb{E}[T]$ *is sufficiently small, in the following sense:*

$$\left|\bar{\Delta}^{w\text{-}peak}(\mathbf{r}^\star) - \bar{\Delta}^{w\text{-}peak}_{opt}\right| \leq \frac{t_s}{\mathbb{E}[T]}C_2 + o\left(\frac{t_s}{\mathbb{E}[T]}\right), \quad (28)$$

where

$$C_2 = \sum_{l=1}^{M} \frac{w_l}{b_l(1 - \sum_{i=1}^{M} b_i)}\left(3\sum_{i=1}^{M} b_i - \min_j b_j\right). \quad (29)$$

Proof. See Appendix C.2. □

We obtain the following corollary from Theorem 3.

Corollary 4 (**Asymptotic optimality**). *If* $\sum_{i=1}^{M} b_i < 1$, *then* (23) *holds for the solution* $\mathbf{r}^\star$ (18) *and* (25) - (27). *In other words, our proposed solution is asymptotically optimal for Problem* ***1*** *in* (17) *as* $t_s/\mathbb{E}[T] \to 0$. *Moreover, the asymptotic optimal objective value of Problem* ***1*** *as* $t_s/\mathbb{E}[T] \to 0$ *is*

$$\begin{aligned}\lim_{\frac{t_s}{\mathbb{E}[T]}\to 0} \bar{\Delta}^{w\text{-}peak}_{opt} &= \sum_{i=1}^{M}\left[\frac{w_i}{\min\{b_i, \beta^\star\sqrt{w_i}\}} + w_i\right] \\ &= \sum_{i=1}^{M}\left[\frac{w_i}{b_i} + w_i\right].\end{aligned} \quad (30)$$

Proof. See Appendix C.2. □

Interestingly, the asymptotic optimal objective values of Problem **1** in both regimes, given by (24) and (30), are of an identical expression. However, in the energy-scarce regime, we can observe that $\beta^\star$, which is defined in (25), always satisfies $\min\{b_l, \beta^\star\sqrt{w_l}\} = b_l$ for all l.

Remark 1. *We would like to point out that the condition* $t_s/\mathbb{E}[T] \approx 0$ *is satisfied in many practical applications. For instance, in a wireless sensor network that is equipped with low-power UHF transceivers [47], the carrier sensing time is* $t_s = 40$ μ*s, while the transmission time is around* 5 *ms. Hence,* $t_s/\mathbb{E}[T] \approx 0.008$.

3.3 Discussion

In this subsection, we present a simple implementation of our proposed solution, discuss the nested convex optimization method that can be used to solve Problem **1** when $t_s = 0$, provide some useful insights about our proposed solution at the limit point $t_s/\mathbb{E}[T] \to 0$, and provide a comparison with synchronized schedulers performance.

3.3.1 Implementation of Sleep-wake Scheduling

We devise a simple algorithm to compute our solution $\mathbf{r}^\star$, which is provided in Algorithm 1. Notice that $\mathbf{r}^\star$ has the

same expression (18) in the energy-adequate and energy-scarce regimes. We exploit this fact to simplify the implementation of sleep-wake scheduling. In particular, the sources report w_l and b_l to the AP, which computes $\beta^\star$ and $x^\star$, and broadcasts them back to the sources. After receiving $\beta^\star$ and $x^\star$, source l computes $r_l^\star$ based on (18). In practical wireless sensor networks, e.g., smart city networks and industrial control sensor networks [48], [49], the sensors report their measurements via an access point (AP). Hence, it is reasonable to employ the AP in implementing the sleep-wake scheduler.

Algorithm 1: Implementation of sleep-wake scheduler.

```
The AP gathers the parameters {(w_i, b_i)_{i=1}^M, t_s/E[T]};
if Σ_{i=1}^M b_i ≥ 1 then
    The AP computes x*, β* from (19) and (20);
else
    The AP computes x*, β* from (25) - (27);
end
The AP broadcasts x*, β* to all the M sources;
Upon hearing x*, β*, source l compute r_l* from (18);
```

In the above implementation procedure, the sources do not need to know if the overall network is in the energy-adequate or energy-scarce regime; only the AP knows about it. Further, the amount of downlink signaling overhead is small, because only two parameters $\beta^\star$ and $x^\star$ are broadcasted to the sources. Moreover, when the node density is high, the scalability of the network is a crucial concern and reporting w_l and b_l for each source is impractical. In this case, the AP can compute $\beta^\star$ and $x^\star$ by estimating the distribution of w_l and b_l, as well as the number of source nodes, which reduces the uplink signaling overhead. Finally, when sources are not in the hearing range of each other, hidden/exposed source problems arise. These problems are challenging to solve analytically. However, this can be solved by designing practical heuristic solutions based on the theoretical solutions. One design method was given in [23].

3.3.2 The Nested Convex Optimization Method for $t_s = 0$

If $t_s = 0$, Problem **1** reduces to the following optimization problem:

$$\bar{\Delta}_{\text{opt}}^{\text{w-peak}} \triangleq \min_{r_l>0} \sum_{l=1}^{M} \frac{w_l \left(1+\sum_{i=1}^{M} r_i\right)}{r_l} + \sum_{l=1}^{M} w_l \quad \text{s.t. } r_l \le b_l(\sum_{i=1}^{M} r_i + 1), \forall l. \tag{31}$$

Observe that the optimization problem in (31) is non-convex. To bypass this difficulty, we use an auxiliary variable $y = \sum_{i=1}^{M} r_i + 1$. Hence, we obtain the following optimization problem for given y:

$$\min_{r_i>0} \sum_{i=1}^{M} \left[\frac{w_i y}{r_i} + w_i\right] \tag{32}$$

$$\textbf{s.t. } r_l \le b_l y, \forall l, \tag{33}$$

$$\sum_{i=1}^{M} r_i + 1 = y. \tag{34}$$

The objective function in (32) is a convex function. Moreover, the constraints in (33) and (34) are affine. Hence, Problem (32) is convex. Exploiting (32), we solve (31) by using a two-layer nested convex optimization method: In the inner layer, we optimize $\mathbf{r}$ for given y. After solving $\mathbf{r}$, we will optimize y in the outer layer. This technique is used in the proof of Lemma 9 in Appendix E, where the reader can find the detailed solution.

3.3.3 Asymptotic Behavior of The Optimal Solution

In the energy-adequate regime, the sleeping period parameter $r_l^\star$ of source l tends to infinity as $t_s/\mathbb{E}[T] \to 0$, while the ratio $r_l^\star/r_i^\star$ between source l and source i is kept as a constant for all l and i. Hence, the sleeping time of the sources tends to zero. Meanwhile, since $t_s/\mathbb{E}[T] \to 0$, the sensing time becomes negligible. The channel access probability of source l in this limit can be computed as

$$\lim_{\frac{t_s}{\mathbb{E}[T]}\to 0} \sigma_l^\star = \min\{b_l, \beta^\star\sqrt{w_l}\}. \tag{35}$$

Because of (20), $\lim_{t_s/\mathbb{E}[T]\to 0} \sum_{i=1}^{M} \sigma_i^\star = 1$. Hence, the channel is occupied by the sources at all time, without any time overhead spent on sensing and sleeping.

On the other hand, in the energy-scarce regime, the sleeping period parameter $r_l^\star$ of source l converges to a constant value when $t_s/\mathbb{E}[T] \to 0$, i.e., we have

$$\lim_{\frac{t_s}{\mathbb{E}[T]}\to 0} r_l^\star = \frac{b_l}{1-\sum_{i=1}^{M} b_i}. \tag{36}$$

Since the cumulative energy is scarce, the sources necessarily need to stay idle for some time in order to meet their target lifetime. Hence, sleep periods are imposed for achieving the optimal trade-off between minimizing AoI and energy consumption.

3.3.4 Comparison with Synchronized Schedulers Performance

We would like to show that the performance of our proposed algorithm is asymptotically no worse than any synchronized (e.g., centralized) scheduler. Consider a scheduler

in which the fraction of time during which source l transmits update packets is equal to a_l, where we have $\mathbf{a} = \{a_l\}_{l=1}^M$ and $\sum_{i=1}^M a_i \leq 1$. In this scheduler, only one source is allowed to access the channel at a time, i.e., there is no collision (this can be achieved either by a deterministic scheduler or by assigning a channel access probability a_l for each source l after each packet transmission)[8]. We can perform an analysis similar to that of Section 2.2, and show that the total weighted average peak age of a synchronized scheduler is given by

$$\sum_{i=1}^{M} \left[\frac{w_i \mathbb{E}[T]}{a_i} + w_i \, \mathbb{E}[T] \right]. \tag{37}$$

Hence, the problem of designing an optimal synchronized scheduler that minimizes the total weighted average peak age under energy constraints can be cast as

$$\bar{\Delta}^{\text{w-peak}}_{\text{opt-s}} \triangleq \min_{a_i>0} \sum_{i=1}^{M} \left[\frac{w_i}{a_i} + w_i \right] \tag{38}$$

$$\textbf{s.t.} \; a_l \leq b_l, \; \forall l, \tag{39}$$

$$\sum_{i=1}^{M} a_i \leq 1, \tag{40}$$

where we have divided the objective function by $\mathbb{E}[T]$. Next, we show that the performance of our proposed algorithm converges to that of the optimal synchronized scheduler when $t_s/\mathbb{E}[T] \to 0$.

Corollary 5. *For any $(w_i, b_i)_{i=1}^M$, we have*

$$\lim_{\frac{t_s}{\mathbb{E}[T]} \to 0} \bar{\Delta}^{\textit{w-peak}}_{\textit{opt}} = \bar{\Delta}^{\textit{w-peak}}_{\textit{opt-s}}. \tag{41}$$

Proof. The proof is provided in Appendix H which is listed at the end before Appendix I as it requires some results from precedent appendixes. □

Synchronized schedulers were recently studied in [15] for the case without energy constraints, i.e., $b_l \geq 1$ for all l. According to Corollary 5, the channel access probability of the synchronized scheduler in [15] is a special case of our solution (35) where $b_l \geq 1$ for all l.

4 Learning to Optimize Age

Note that the optimal rate $\mathbf{r}^\star$ in Theorem 1 depends upon the mean transmission time $\mathbb{E}[T]$. Since the transmission time also depends upon (possibly) time-varying channel conditions, estimating $\mathbb{E}[T]$ accurately a priori, could be cumbersome. Thus, in this section, we derive learning algorithms that optimize the total weighted average peak age of all sources when the mean transmission time $\mathbb{E}[T]$ is unknown to the scheduler. We begin by reducing our system to an equivalent discrete-time Markov chain.

Contributions and Challenges: The simplest learning algorithm is called the certainty equivalent (CE) rule [50]–[53]. In this, the scheduler maintains an empirical estimate of $\mathbb{E}[T]$, and utilizes sleep parameters that are optimal when the true value of the mean transmission time is equal to this estimate. The regret of a learning algorithm is the sub-optimality in the performance that results because the algorithm does not know the system parameters. What we are able to show is that by using the CE rule, we are able to get $o(H)$ regret, where H is the time-horizon. This further implies that the *long-term time-average performance* of our CE algorithm is asymptotically optimal.

This result is important since it is well-known by now [54] that in many reinforcement learning problems [40], the CE rule fails to yield long-term time average performance, because it does not yield a correct estimate of the optimal choices. Thus, more complex learning rules, such as optimism in the face of uncertainty [41], [53] that utilize confidence balls in addition to the empirical estimates and thus have a significantly higher computational complexity, are required in order to ensure optimality. Our main contribution is to show that the vanilla CE rule yields asymptotically the same long-term time average performance as the scheduler that knows the system parameters in advance, i.e. (with a high probability) the "sub-optimality gap" of the CE rule is $o(H)$ where H is the operating time horizon. This means that instead of using more complex learning algorithm such as the UCRL [41] or RBMLE [53], one could use CE thereby saving precious computing power and attaining the optimal average performance (asymptotically). We perform a finite-time performance analysis of the CE rule and explicitly quantify its sub-optimality by deriving an upper-bound on its "regret", i.e., the gap between its average expected performance, and that resulting from the application of optimal sleep parameter. The problem of designing and analyzing learning algorithms for our setup poses several challenges, primarily because the age process evolves in continuous time on a continuous state-space that is not compact. To address this difficulty, we show that for the purpose of optimizing average age, we can equivalently work with a discrete-time process. We then utilize several techniques from the theory of general state-space Markov chains [55] for analyzing the learning regret.

Sampling Continuous Time Process: Consider the multi-

8. Note that if $\sum_{i=1}^M a_i < 1$, then it is possible that the scheduler decides not to serve any source after the transmission of some packet. In this case, the scheduler waits for a random time that has the same distribution as the transmission time T before deciding to serve another source.

source system in which the sleep durations are modulated according to the parameter vector $\mathbf{r} = (r_1, r_2, \ldots, r_M)$. Throughout this section, we let $n \in \mathbb{N}$ be the discrete time of the sampled system. We sample the original continuous-time system at those time instants when one out of the following events occur:

- a source l gets channel access and starts transmitting. We say that it wakes up, denoted by $m_l(n) = 1$,
- a source l completes packet transmission, and hence goes into sleep mode such that $m_l(n) = 0$.

In what follows, we make this assumption.

Assumption 1. *The transmission times are bounded, i.e.,* $0 \leq T \leq T_{\max}$ *almost surely, where* $T_{\max} > 0$*. Moreover, the probability density function* $f(\cdot)$ *of* T *satisfies*

$$lb \leq f(y) \leq ub, \forall y \in [0, T_{\max}],$$

where $lb, ub > 0$ *are upper and lower bounds on the density function.* □

Define $s_l(n) := (\Delta_l(n), m_l(n))$, where $\Delta_l(n)$ is the age, and $m_l(n) \in \{0, 1\}$ is the mode of user l. Define,

$$\mathbf{s}(n) := (s_1(n), s_2(n), \ldots, s_M(n)). \tag{42}$$

As is shown in Lemma 15 (see Appendix I), for the purpose of adaptively choosing sleep parameters, the process $\mathbf{s}(n)$ serves as a sufficient-statistics [56] for the optimization problem (17). In other words, $\mathbf{s}(n)$ is the state of a Markov decision process. Hence, we will work exclusively with the discrete-time system obtained by sampling the original continuous-time system. We use $\mathcal{S}$ to denote the state space of a single source, i.e., we have $s_l(n) \in \mathcal{S}$. Consider the operation over a time horizon of H discrete time-steps, and let K_l denote the (random) number of packets delivered to source l until time H. The cumulative cost incurred is given by

$$C(H) := \sum_{l=1}^{M} \sum_{i=1}^{K_l} w_l \Delta_{l,i}^{\text{peak}}, \tag{43}$$

where $\Delta_{l,i}^{\text{peak}}$ denotes the i-th peak age of source l. We let $r_l(n) \in \mathbb{R}_+$ denote the sleep period parameter for source l, and denote $\mathbf{r}(n) := (r_1(n), r_2(n), \ldots, r_M(n))$. As is shown in Lemma 15 in the supplementary material, the expected value of the cumulative value of peak age can be written as follows,

$$\mathbb{E}\left(\sum_{n=1}^{H-1} g(\mathbf{s}(n))\right), \tag{44}$$

where the function g is described in Lemma 15. However, in our setup, the controller that chooses $\mathbf{r}(n)$ does not know the density function f of the packet transmission time, and has to adaptively choose the sleeping period paremeter $\mathbf{r}(n)$ so as to minimize the operating cost (44).

A learning policy is a rule that chooses the sleep period parameter adaptively based on the past operation history of the system, i.e. the values of the random variables $\{\mathbf{s}(i)\}_{i=1}^{n}, \{\mathbf{r}(i)\}_{i=1}^{n-1}$. Throughout this section we use θ to denote the mean transmission time $\mathbb{E}[T]$. Since the optimal rate depends upon the probability density function $f(\cdot)$ of the transmission time T only through its mean $\theta = \mathbb{E}[T]$, we also denote it by $\mathbf{r}_\theta^\star$. The performance of a learning policy is measured by its regret $R(H)$, which is defined by

$$R(H) := \sum_{n=1}^{H} g(\mathbf{s}(n)) - H\bar{\Delta}^{\text{w-peak}}(\mathbf{r}_\theta^\star), \tag{45}$$

where $\bar{\Delta}^{\text{w-peak}}(\mathbf{r}_\theta^\star)$ is used to denote the optimal age performance in (17) when the true system parameter is known and hence the scheduler can implement the optimal rate vector.

Certainty Equivalence Learning Algorithm: We begin with some notations. Let $col(i), i = 1, 2, \ldots$ be a random variable that is equal to 1 if there is no collision at time i, and is 0 otherwise. Define $N(n) := \sum_{i=1}^{n} col(i)$ as the number of packet transmissions without collisions by time n. The empirical estimate of $\theta = \mathbb{E}[T]$ at time n is denoted by $\hat{\theta}(n)$, and given as

$$\hat{\theta}(n) := \frac{\sum_{i=1}^{n} T(i) col(i)}{\max\{N(n), 1\}}, \tag{46}$$

where $T(i) \in [0, T_{\max}]$ is the time taken to deliver packet at time i.

The learning rule operates in episodes. We let τ_k be the start time of the k-th episode, and let $\mathcal{E}_k := \{\tau_k, \tau_k + 1, \ldots, \tau_{k+1} - 1\}$ be the time-slots that comprise the k-th episode, so that the duration of $\mathcal{E}_k$ is $\tau_{k+1} - \tau_k$ time-slots. We let $\tau_k = 2^k$, use $k(n)$ to denote the index of the current episode at time n, and $\theta(n)$ to denote the empirical estimate at the beginning of the current episode, defined by

$$k(n) := \max\{k : \tau_k \leq t\}, \tag{47}$$

$$\theta(n) := \hat{\theta}(\tau_{k(n)}). \tag{48}$$

Within each single episode the algorithm implements a single stationary controller that makes decisions only on the basis of the state $\mathbf{s}(n)$ and the estimate $\theta(\tau_k)$ obtained at the beginning of the current ongoing episode $k(n)$. It chooses the sleep period parameter as $\mathbf{r}(n) = \mathbf{r}_{\theta(n)}^\star, \forall n \in \mathcal{E}_k$, i.e., it utilizes the rate vector that is optimal for the system whose

mean transmission time is equal to $\theta(n)$. Thus, $\mathbf{r}(n) = \mathbf{r}^{\star}_{\theta(\tau_k)}$ for $\tau_k \leq n \leq \tau_{k+1} - 1$.

We summarize our learning rule in Algorithm 2. We will analyze its performance under the following assumptions. Throughout, for a vector $\mathbf{x}$, we let $\|\mathbf{x}\|$ denote its Euclidean norm, and $\|\mathbf{x}\|_1$ denote its 1-norm.

Algorithm 2: Certainty Equivalence Learning for Age Optimization

Input: $N, \gamma \geq 4$
Set $\hat{\theta}(1) = .5$.
1: **for** $n = 1, 2, \ldots$ **do**
2: **if** $n = \tau_k$ **then**
3: Calculate $\hat{\theta}(n)$ as in (46) and set $\theta(n)$ as in (47)-(48).
4: **end if**
Compute $\mathbf{r}(n) = \mathbf{r}^{\star}_{\theta(n)}$ based on the results in Section 3.
5: **end for**

Assumption 2. *With a high probability, say greater than $1 - \delta$, where $\delta > 0$ is a small constant, the state value $\mathbf{s}(\tau_k)$ at the beginning of each episode k belongs to a compact set $\mathcal{K} := \left\{\mathbf{x} \in \mathcal{S}^M : \|\mathbf{x}\|_1 \leq K_1\right\}$, where $\mathcal{S}$ is the state space of a single source.* □

The above is not a restrictive assumption, since the scheduler can always ensure that towards the end of each episode, each source receives a sufficient amount of service in order to ensure this condition. Let Θ denote the set of allowable values of the mean transmission time $\theta = \mathbb{E}[T]$. We now make an assumption regarding the set Θ.

Assumption 3. *Recall that $\mathbf{r}^{\star}_{\theta}$ is the optimal sleep parameter when the mean transmission time is equal to θ. The following two properties hold for the scheduler that uses $\mathbf{r}(n) \equiv \mathbf{r}^{\star}_{\theta}, n \in \mathbb{N}$.*
(i) The average cost is finite, i.e.,

$$\limsup_{H\to\infty} \frac{1}{H} \sum_{n=1}^{H} \mathbb{E}_{\mathbf{r}^{\star}_{\theta}} \left(g(\mathbf{s}(n))\right) \leq K_2 < \infty, \tag{49}$$

holds for all $\theta \in \Theta$.
(ii) Each user gets channel access with a non-zero probability

$$\inf_{\theta\in\Theta, l\in[M]} \mathbb{P}\left(ca_l(n) = 1 | \mathbf{r}(n) = \mathbf{r}^{\star}_{\theta}\right) > 0, \tag{50}$$

where $ca_l(i)$ is a random variable that is 1 if source l gets channel access at time i, while is 0 otherwise. We denote

$$p_{\min} := \inf_{\theta\in\Theta, l\in[M]} \mathbb{P}\left(ca_l(n) = 1 | \mathbf{r}(n) = \mathbf{r}^{\star}_{\theta}\right). \tag{51}$$

□

It is easily verified that (49), (50) hold true for a system in which the rate vector $\mathbf{r}$ is bounded, since each source gets access to the channel with non zero probability within a finite time.

The following result quantifies the learning regret of Algorithm 2.

Theorem 6. *Consider the problem of designing a learning algorithm that does not know the statistics of the transmission time T, and adaptively chooses the sleep period parameters $\mathbf{r}(n)$ in order to minimize the cumulative peak age of M sources. Let $\delta_1 \in (0, p_{\min})$ be a constant. Then, under Assumptions 1-3, the regret of Algorithm 2 can be bounded as follows,*

$$\mathbb{E}\left[R(H)\right] \leq K_2 \max\left\{\frac{\gamma \log H}{(p_{\min} - \sqrt{\delta_1})\delta^2}, O\left(\frac{1}{\delta_1}\log H\right)\right\} + K_2\frac{\pi^2}{6} + L\sqrt{\frac{H\gamma(\log H)^2}{(p_{\min} - \sqrt{\delta_1})}}.$$

where H is the operating time horizon, $\gamma \geq 4$ is a constant, $K_2, p_{\min}$ are as in Assumption 3, and the parameters $\delta, L > 0$ are as in Lemma 19.

Proof. See Appendix I. □

5 Numerical and Simulation Results

We use Matlab and NS-3 to evaluate the performance of our algorithm. We use "age-optimal scheduler" to denote the sleep-wake scheduler with the sleep period paramters $r^{\star}_l$'s as in (18), which was shown to be near-optimal in Theorem 1 and Theorem 3. By "throughput-optimal scheduler", we refer to the sleep-wake algorithm of [23] that is known to achieve the optimal trade-off between the throughput and energy consumption reduction. Moreover, we use "fixed sleep-rate scheduler" to denote the sleep-wake scheduler in which the sleep period parameters r_l's are equal for all the sources, i.e., $r_l = k$ for all l, where the parameter k has been chosen so as to satisfy the energy constraints of Problem **1**. We also let $\bar{\Delta}^{\text{w-peak}}_{\text{un}}(\mathbf{r})$ denote the unnormalized total weighted average peak age in (11). Finally, we would like to mention that we do not compare the performance of our proposed algorithm with the CSMA algorithms of [36], [37] where the goal was solely to minimize the age. Since they do not incorporate energy constraints, it is not fair to compare the performance of our algorithm with them.

Unless stated otherwise, our set up is as follows: The average transmission time is $\mathbb{E}[T] = 5$ ms. The weights w_l's attached to different sources are generated by sampling from a uniform distribution in the interval $[0, 10]$. The target power efficiencies b_l's are randomly generated according to a uniform distribution in the range $[0, 1]$.

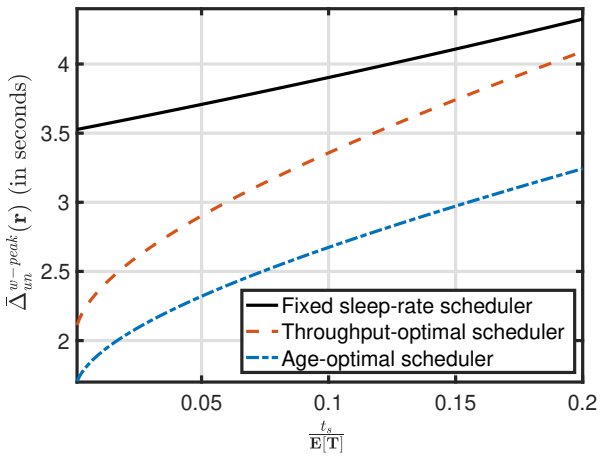

Figure 3: Total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) versus the ratio $\frac{t_s}{\mathbb{E}[T]}$ for $M = 10$ sources.

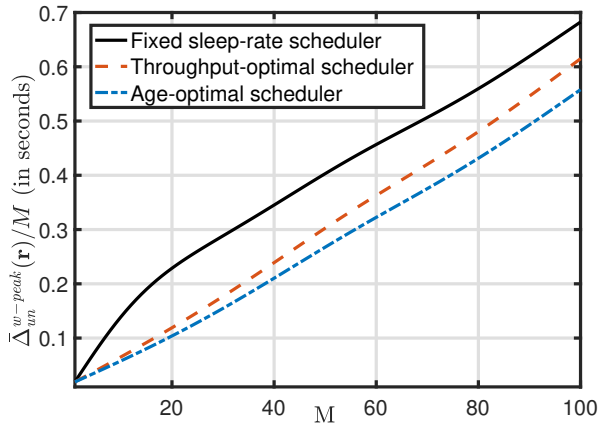

Figure 4: Total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) versus the number of sources M, where $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ has been normalized by M while plotting.

5.1 Numerical Evaluations

Figure 3 plots the total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) as a function of the ratio $\frac{t_s}{\mathbb{E}[T]}$, where the number of sources is $M = 10$. The age-optimal scheduler is seen to outperform the throughput-optimal and Fixed sleep-rate schedulers. This implies that what minimizes the throughput does not necessarily minimize AoI and vice versa. Moreover, we observe that the total weighted average peak age of all schedulers increases as the sensing time increases. This is expected since an increase in the sensing time leads to an increase in the probability of packet collisions, which in turn deteriorates the age performance of these schedulers.

We then scale the number of sources M, and plot $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) as a function of M in Figure 4. While plotting, we normalize the performance by the number of sources M. The sensing time t_s is fixed at $t_s = 40$ μs.

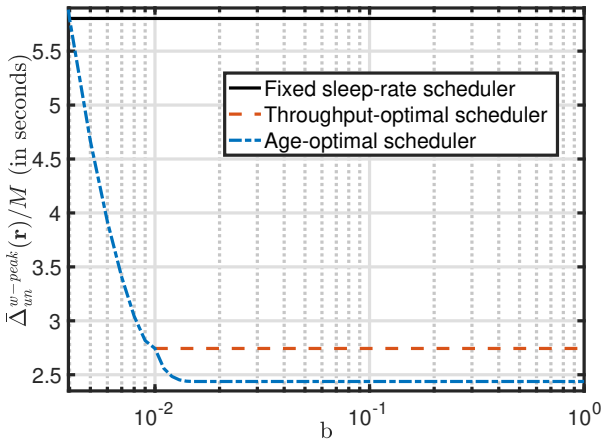

Figure 5: Total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) versus the target power efficiency b for $M = 100$ sources, where $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ has been normalized by M while plotting.

The weights w_l's corresponding to different sources are randomly generated uniformly within the range $[0, 2]$. The age-optimal scheduler is shown to outperform other schedulers uniformly for all values of M. Moreover, as we can observe, the average peak age of the sources under age-optimal scheduler increases up to around 0.55 seconds only, while the number of sources rises from 1 to 100. This indicates the robustness of our algorithm to changes in the number of sources in a network.

In Figure 5, we fix the value of M as 100 and the target power efficiencies at the same value for all the sources, i.e., $b_l = b$ for all l. We then vary the parameter b and plot the resulting performance. While plotting, we normalize the performance by the number of sources M. We exclude the simulation of the throughput-optimal scheduler for $b < 0.01$ since the sleeping period parameters that are proposed in [23] are not feasible for Problem **1** in the energy-scarce regime, i.e., when $\sum_{i=1}^{M} b_i < 1$. The age-optimal scheduler outperforms the other schedulers. Moreover, its performance is a decreasing function of b, and then settles at a constant value. This occurs because our proposed solution in (18) is a function solely of the weights w_l's and $\beta^\star$ when b exceeds some value. Thus, the performance of the proposed scheduler saturates after this value of b.

We now show the effectiveness of the proposed scheduler when deployed in "dense networks" [21], [22]. Dense networks are characterized by a large number of sources connected to a single AP. We fix M at 10^5 sources, and take the target lifetimes of the sources to be equal, i.e., $D_l = D$ for all l. The weights w_l's corresponding to different sources are generated randomly by sampling from the uniform distribution in the range $[0, 2]$. We let the initial battery

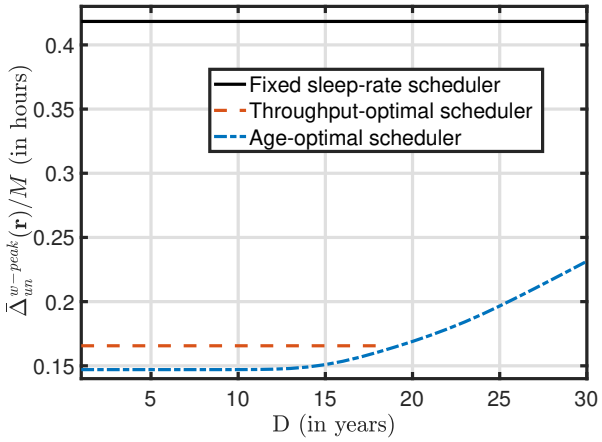

Figure 6: Total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ in (11) versus the target lifetime D for a dense network with $M = 10^5$ sources, where $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ has been normalized by M while plotting. Since the throughput–optimal scheduler is infeasible for values of D greater than 18 years, we do not plot its performance for these values.

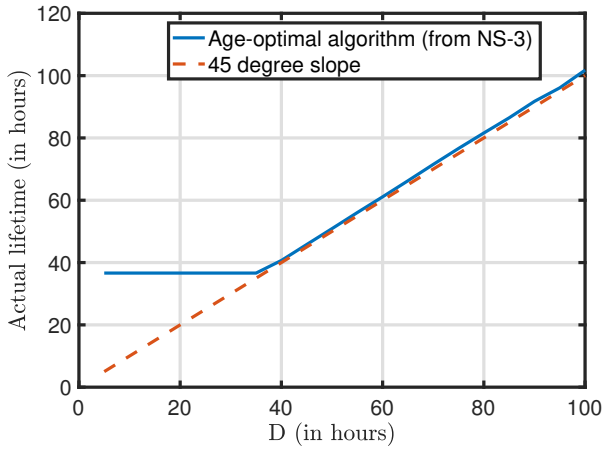

Figure 7: The average actual lifetime versus the target lifetime D.

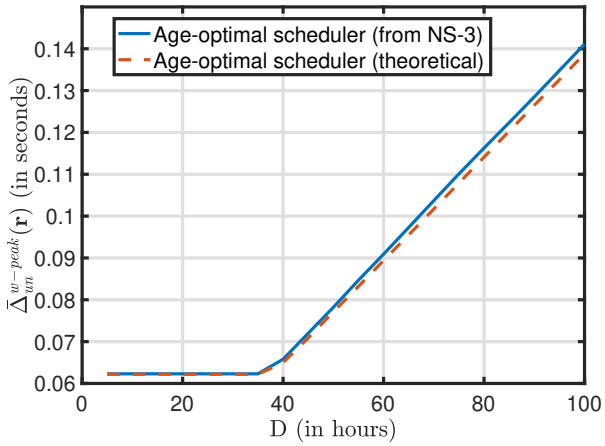

Figure 8: Total weighted average peak age $\bar{\Delta}_{\text{un}}^{\text{w-peak}}(\mathbf{r})$ versus the target lifetime D.

level $B_l = 8$ mAh for all l and the output voltage is 5 Volt. We also let the energy consumption in a transmission mode to be 24.75 mW for all sources. We vary the parameter D and plot the resulting performance in Figure 6. While plotting, we normalize the performance by the number of sources M. We exclude simulations for the throughput-optimal scheduler for values of D for which the scheduler is infeasible, i.e., its cumulative energy consumption exceeds the total allowable energy consumption. The age-optimal scheduler is seen to outperform the others. As observed in Figure 6, under the age-optimal scheduler, sources can be active for up to 25 years, while simultaneously achieving a decent average peak age of around .2 hour, i.e., 12 minutes. This makes it suitable for dense networks, where it is crucial that the sources are necessarily active for many years.

5.2 NS-3 Simulation

We use NS-3 [57] to investigate the effect of our model assumptions on the performance of age-optimal scheduler in a more practical situation. We simulate the age-optimal scheduler by using IEEE 802.11b while disabling the RTS-CTS and modifying the back-off times to be exponentially distributed in the MAC layer. Our simulation results are averaged over 5 system realizations. The UDP saturation conditions are satisfied such that the source nodes always have packets to send.

Our simulation consists of a WiFi network with 1 AP and 3 associated source nodes in a field of size 50m $\times$ 50m. We set the sensing threshold to -100 dBm which covers a range of 110m. Thus, all sources can hear each other. The initial battery level of each source is 60 mAh, where the output voltage is 5 Volt. For each source, the power consumption in the transmission mode is 24.75 mW, and the power consumption in the sleep mode is 15 μW. Moreover, all weights are set to unity, i.e., $w_l = 1$ for all l.

Figure 7 plots the average actual lifetime of the sources versus the target lifetime, where we take the target lifetimes of all sources to be equal, i.e., $D_l = D$ for all l. As we can observe, the actual lifetime of the age-optimal scheduler always achieves the target lifetime. This suggests that our assumptions (i.e., (i) omitting the power dissipation in the sleep mode and in the sensing times, (ii) the average transmission times and collision times are equal to each other) do not affect the performance of the algorithm which reaches its target lifetime.

Figure 8 plots the total weighted average peak age versus the target lifetime, where again we take the target lifetimes of all sources to be equal, i.e., $D_l = D$ for all l. The

age-optimal scheduler (theoretical) curve is obtained using (11), while the age-optimal scheduler (from NS-3) curve is obtained using the NS-3 simulator. As we can observe, the difference between the plotted curves does not exceed 2% of the age-optimal scheduler (theoretical) performance. This emphasizes the negligible impact of our assumptions on the performance of our proposed algorithm.

6 Conclusions

We designed an efficient sleep-wake scheduling algorithm for wireless networks that attains the optimal trade-off between minimizing the AoI and energy consumption. Since the associated optimization problem is non-convex, in general we could not hope to solve it for all values of the system parameters. However, in the regime when the carrier sensing time t_s is negligible as compared to the average transmission time $\mathbb{E}[T]$, we were able to provide a near-optimal solution. Moreover, the proposed solution is in a simple form that allowed us to design an easy-to-implement algorithm to obtain the solution. Furthermore, we showed that the performance of our proposed algorithm is asymptotically no worse than that of the optimal synchronized scheduler, as $t_s/\mathbb{E}[T] \to 0$. Finally, when the mean transmission time is unknown, we devise a reinforcement learning algorithm that adaptively learns the unknown parameter.

7 Acknowledgements

The authors appreciate Jiayu Pan and Shaoyi Li for their great efforts in obtaining the ns-3 simulation results.

References

[1] A. M. Bedewy, Y. Sun, R. Singh, and N. B. Shroff, "Optimizing information freshness using low-power status updates via sleep-wake scheduling," in *Proc. MobiHoc*, 2020, pp. 51–60.

[2] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE INFOCOM*, 2012, pp. 2731–2735.

[3] Y. Sun, I. Kadota, R. Talak, and E. Modiano, "Age of information: A new metric for information freshness," *Synthesis Lectures on Communication Networks*, vol. 12, no. 2, pp. 1–224, 2019.

[4] R. D. Yates, Y. Sun, D. R. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *arXiv preprint arXiv:2007.08564*, 2020.

[5] N. F. Timmons and W. G. Scanlon, "Analysis of the performance of ieee 802.15. 4 for medical sensor body area networking," in *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks. IEEE SECON 2004.*, 2004, pp. 16–24.

[6] Chipcon AS SmartRF CC3420 Preliminary Datasheet, rev 1.0, 17 November 2003.

[7] Datasheet for MC9S08RE8 motorola microcontroller.

[8] R. D. Yates and S. K. Kaul, "Status updates over unreliable multiaccess channels," in *Proc. IEEE ISIT*, 2017, pp. 331–335.

[9] R. Talak, S. Karaman, and E. Modiano, "Distributed scheduling algorithms for optimizing information freshness in wireless networks," in *Proc. IEEE SPAWC*, 2018, pp. 1–5.

[10] R. Li, A. Eryilmaz, and B. Li, "Throughput-optimal wireless scheduling with regulated inter-service times," in *Proc. IEEE INFOCOM*, 2013, pp. 2616–2624.

[11] 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 Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, 2018.

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

[13] Z. Jiang, B. Krishnamachari, X. Zheng, S. Zhou, and Z. Niu, "Timely status update in massive iot systems: Decentralized scheduling for wireless uplinks," *arXiv preprint arXiv:1801.03975*, 2018.

[14] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. INFOCOM*, 2018, pp. 1844–1852.

[15] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. MobiHoc*, 2018, pp. 61–70.

[16] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, 2017.

[17] X. Guo, R. Singh, P. R. Kumar, and Z. Niu, "A risk-sensitive approach for packet inter-delivery time optimization in networked cyber-physical systems," *IEEE/ACM Trans. Netw.*, vol. 26, no. 4, pp. 1976–1989, 2018.

[18] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *IEEE INFOCOM - the 1st Workshop on the Age of Information (AoI Workshop)*, 2018, pp. 136–141.

[19] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Proc. Allerton*, 2016, pp. 844–851.

[20] R. Singh, X. Guo, and P. R. Kumar, "Index policies for optimal mean-variance trade-off of inter-delivery times in real-time sensor networks," in *Proc. IEEE INFOCOM*. IEEE, 2015, pp. 505–512.

[21] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient coded random access for the wireless uplink," *arXiv preprint arXiv:1907.09448*, 2019.

[22] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user mac with finite payloads and fading," *arXiv preprint arXiv:1901.06732*, 2019.

[23] S. Chen, T. Bansal, Y. Sun, P. Sinha, and N. B. Shroff, "Life-add: Lifetime adjustable design for wifi networks with heterogeneous energy supplies," in *Proc. WiOpt*, 2013, pp. 508–515.

[24] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2018.

[25] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, 2016.

[26] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. IEEE ISIT*, 2016, pp. 2569–2573.

[27] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5215–5232, 2019.

[28] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, 2017, pp. 576–580.

[29] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, 2019.

[30] 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, no. 11, pp. 7492–7508, 2017.

[31] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 204–219, 2019.

[32] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in *Proc. MobiHoc*, 2019, pp. 121–130.

[33] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Optimal sampling and scheduling for timely status updates in multi-source networks," *IEEE Trans. Inf. Theory*, pp. 1–1, 2021.

[34] S. Yun, Y. Yi, J. Shin, et al., "Optimal CSMA: a survey," in *Proc. ICCS*, 2012, pp. 199–204.

[35] R. Singh and P. R. Kumar, "Adaptive CSMA for decentralized scheduling of multi-hop networks with end-to-end deadline constraints," *IEEE/ACM Trans. Netw.*, pp. 1–14, 2021.

[36] A. Maatouk, M. Assaad, and A. Ephremides, "Minimizing the age of information in a CSMA environment," *arXiv preprint arXiv:1901.00481*, 2019.

[37] M. Wang and Y. Dong, "Broadcast age of information in CSMA/CA based wireless networks," *arXiv preprint arXiv:1904.03477*, 2019.

[38] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, 2010.

[39] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.

[40] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, MIT press, 1998.

[41] T. Jaksch, R. Ortner, and P. Auer, "Near-optimal regret bounds for reinforcement learning.," *Journal of Machine Learning Research*, vol. 11, no. 4, 2010.

[42] R. Singh, A. Gupta, and N. B. Shroff, "Learning in Markov decision processes under constraints," *arXiv preprint arXiv:2002.12435*, 2020.

[43] A. N. Shiryaev, *Optimal stopping rules*, New York: Springer-Verlag, 1978.

[44] A. Wald, *Sequential analysis*, New York: Courier Corporation, 1973.

[45] ASH transceiver TR1000 data sheet, RF Monolithic Inc.

[46] K. F. Ramadan, M. I. Dessouky, M. Abd-Elnaby, and F. E. A. El-Samie, "Energy-efficient dual-layer MAC protocol with adaptive layer duration for wsns," in *11th International Conference on Computer Engineering Systems (ICCES)*, 2016, pp. 47–52.

[47] A. El-Hoiydi, "Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks," in *Proc. IEEE Int. Symp. Comput. Commun. (ISCC)*, 2002, pp. 685–692.

[48] C. Lu, A. Saifullah, B. Li, M. Sha, H. Gonzalez, D. Gunatilaka, C. Wu, L. Nie, and Y. Chen, "Real-time wireless sensor-actuator networks for industrial cyber-physical systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1013–1024, 2016.

[49] P. Hsieh and I. Hou, "A decentralized medium access protocol for real-time wireless ad hoc networks with unreliable transmissions," in *IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 972–982.

[50] H. Van de Water and J. Willems, "The certainty equivalence property in stochastic control theory," *IEEE Transactions on Automatic Control*, vol. 26, no. 5, pp. 1080–1087, 1981.

[51] H. Mania, S. Tu, and B. Recht, "Certainty equivalence is efficient for linear quadratic control," in *Advances in Neural Information Processing Systems*, 2019, pp. 10154–10164.

[52] P. Mandl, "Estimation and control in Markov chains," *Advances in Applied Probability*, pp. 40–60, 1974.

[53] A. Mete, R. Singh, X. Liu, and P. R. Kumar, "Reward biased maximum likelihood estimation for reinforcement learning," *Learning for Dynamics and Control (L4DC)*, 2021.

[54] V. Borkar and P. Varaiya, "Adaptive control of Markov chains, i: Finite parameter set," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 953–957, 1979.

[55] E. Nummelin, *General irreducible Markov chains and non-negative operators*, vol. 83, Cambridge University Press, 2004.

[56] C. Striebel, "Sufficient statistics in the optimum control of stochastic systems," *Journal of Mathematical Analysis and Applications*, vol. 12, no. 3, pp. 576–592, 1965.

[57] "NS-3," https://www.nsnam.org/.

[58] R. G. Gallager, *Discrete stochastic processes*, Boston: Kluwer Academic Publishers, 1996.

[59] S. Boyd and L. Vandenberghe, *Convex optimization*, New York, NY, USA: Cambridge University Press, 2004.

[60] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer Science & Business Media, 2012.

[61] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming (Wiley Series in Probability and Statistics)*, Wiley-Interscience, 2005.

[62] P. Billingsley, *Probability and measure*, John Wiley & Sons, 2008.

[63] O. Hernández-Lerma and J. B. Lasserre, *Further topics on discrete-time Markov control processes*, vol. 42, Springer Science & Business Media, 2012.

[64] J. K. Hunter, "Introduction to applied mathematics," lecture notes of Math 207A, University of California Davis, CA, Fall Quarter, 2011.

[65] W. Hoeffding, "Probability inequalities for sums of bounded random variables," in *The Collected Works of Wassily Hoeffding*, pp. 409–426. Springer, 1994.

[66] R. Ayoub, "Euler and the zeta function," *The American Mathematical Monthly*, vol. 81, no. 10, pp. 1067–1086, 1974.

Ahmed M. Bedewy received the B.S. and M.S. degrees in electrical and electronics engineering from Alexandria University, Alexandria, Egypt, in 2011 and 2015, respectively. He received the Ph.D. degree in the Electrical and Computer Engineering from the Ohio State University, OH, USA in 2021. His research interests include wireless communication, cognitive radios, resource allocation, communication networks, information freshness, optimization, and scheduling algorithms. He received the Awarded Certificate of Merit, First Class Honors, for being one of the top ten undergraduate students during 2006-2008 and **1-st** during 2008-2011 in electrical and electronics engineering. His article received the runner-up for the Best Paper Award of ACM MobiHoc 2020.

Yin Sun (S'08-M'11-SM'20) is an Assistant Professor in the Department of Electrical and Computer Engineering at Auburn University, Alabama. He received the B.Eng. and Ph.D. degrees in Electronic Engineering from Tsinghua University, in 2006 and 2011, respectively. He was a Postdoctoral Scholar and Research Associate at the Ohio State University from 2011-2017. His research interests include Age of Information, Networking, Robotic Control, Information Theory, and Machine Learning. He is a senior member of the IEEE and a member of the ACM. He co-founded the Age of Information Workshop in 2018. His articles received the Best Student Paper Award from the IEEE/IFIP WiOpt 2013, Best Paper Award from the IEEE/IFIP WiOpt 2019, runner-up for the Best Paper Award of ACM MobiHoc 2020, and 2021 Journal of Communications and Networks (JCN) Best Paper Award. He co-authored a monograph *Age of Information: A New Metric for Information Freshness*, which was published by Morgan & Claypool Publishers in 2019. He received the Auburn Author Award of 2020.

Rahul Singh Rahul Singh received the B. Tech. degree in Electrical Engineering from Indian Institute of Technology, Kanpur, India in 2009, M.S. degree in Electrical Engineering from the University of Notre Dame, South Bend, USA in 2011, and the Ph.D. degree from Texas A& M University, College Station, in 2015. Currently, he is an Assistant Professor at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. His research interests include stochastic control, machine learning, decentralized control of large-scale complex cyberphysical systems, and scheduling of networks serving real time traffic. His article was runner-up for the Best Paper Award of ACM MobiHoc 2020.

Ness B. Shroff (S'91–M'93–SM'01–F'07) received the Ph.D. degree in electrical engineering from Columbia University in 1994. He joined Purdue University immediately thereafter as an Assistant Professor with the School of Electrical and Computer Engineering. At Purdue, he became a Full Professor of ECE and the director of a university-wide center on wireless systems and applications in 2004. In 2007, he joined The Ohio State University, where he holds the Ohio Eminent Scholar Endowed Chair in networking and communications, in the departments of ECE and CSE. He holds or has held visiting (chaired) professor positions at Tsinghua University, Beijing, China, Shanghai Jiaotong University, Shanghai, China, and IIT Bombay, Mumbai, India. He has received numerous best paper awards for his research and is listed in Thomson Reuters' on The World's Most Influential Scientific Minds, and is noted as a Highly Cited Researcher by Thomson Reuters. He also received the IEEE INFOCOM Achievement Award for seminal contributions to scheduling and resource allocation in wireless networks. He currently serves as the steering committee chair for ACM Mobihoc and the Editor in Chief of the IEEE/ACM Transactions on Networking.

8 Appendix

Appendix A
Derivation of (5)

Define S_l as the residual sleeping period of source l after a sleep-wake cycle is over. Due to the memoryless property of exponential distribution, since the sleeping period of source l is exponentially distributed with mean value $\mathbb{E}[T]/r_l$, S_l is also exponentially distributed with mean value $\mathbb{E}[T]/r_l$. According to the proposed sleep-wake scheduler, source l gains access to the channel and transmits successfully in a given cycle if $S_i \geq S_l + t_s$ for all $i \neq l$. Hence, we have

$$\alpha_l = \mathbb{P}(S_i \geq S_l + t_s,\ \forall i \neq l) \tag{52}$$
$$\overset{(a)}{=} \mathbb{E}[\mathbb{P}(S_i \geq S_l + t_s,\ \forall i \neq l | S_l)] \tag{53}$$
$$\overset{(b)}{=} \mathbb{E}\left[\prod_{i\neq l} \mathbb{P}(S_i \geq S_l + t_s | S_l)\right] \tag{54}$$
$$= \int_0^\infty \left[\prod_{i\neq l} e^{-r_i \frac{s_l + t_s}{\mathbb{E}[T]}}\right] \frac{r_l}{\mathbb{E}[T]} e^{-r_l \frac{s_l}{\mathbb{E}[T]}} ds_l \tag{55}$$
$$= \frac{r_l e^{r_l \frac{t_s}{\mathbb{E}[T]}}}{e^{\sum_{i=1}^M r_i \frac{t_s}{\mathbb{E}[T]}} \sum_{i=1}^M r_i}, \tag{56}$$

where (a) is due to $\mathbb{P}[A] = \mathbb{E}[\mathbb{P}(A|B)]$, and (b) is due to the fact that S_l is independent for different sources. □

Appendix B
Derivation of (13)

Recall the definition of S_l at the beginning of Appendix A. Moreover, define P_l as the probability that source l transmits a packet in a given cycle, regardless whether packet collision occurs or not. For the sleep-wake scheduling mechanism the we are utilizing here, source l transmits in a given cycle as long as no other source wakes up before $S_l - t_s$, i.e., $S_i \geq S_l - t_s$ for all $i \neq l$. Hence, we have

$$P_l = \mathbb{P}(S_i \geq S_l - t_s,\ \forall i \neq l) \tag{57}$$
$$= \mathbb{P}(S_i \geq S_l - t_s,\ \forall i \neq l,\ S_l \geq t_s) + \mathbb{P}(S_l < t_s), \tag{58}$$

where the first term in the RHS is given by

$$\mathbb{P}(S_i \geq S_l - t_s \geq 0,\ \forall i \neq l) \tag{59}$$
$$= \mathbb{E}[\mathbb{P}(S_i \geq S_l - t_s \geq 0,\ \forall i \neq l | S_l)] \tag{60}$$
$$= \mathbb{E}\left[\prod_{i\neq l} \mathbb{P}(S_i \geq S_l - t_s \geq 0 | S_l)]\right] \tag{61}$$
$$= \int_{t_s}^\infty \left[\prod_{i\neq l} e^{-r_i \frac{s_l - t_s}{\mathbb{E}[T]}}\right] \frac{r_l}{\mathbb{E}[T]} e^{-r_l \frac{s_l}{\mathbb{E}[T]}} ds_l \tag{62}$$
$$= e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \frac{r_l}{\sum_{i=1}^M r_i}. \tag{63}$$

Since S_l is exponentially distributed with mean value $\mathbb{E}[T]/r_l$, we can determine the second term in the RHS of (58) as follows:

$$\mathbb{P}(S_l < t_s) = 1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}. \tag{64}$$

Substituting (63) and (64) back into (58), we get

$$P_l = 1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}} + e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \frac{r_l}{\sum_{i=1}^M r_i}. \tag{65}$$

Let α_{col} denote the collision probability in a given cycle. We have $\alpha_{\text{col}} = 1 - \sum_{i=1}^M \alpha_i$, because each cycle includes either a successful transmission or a collision. Moreover, let $\mathbb{E}[\mathbf{Idle}]$ denote the mean of the idle duration in a cycle. By the renewal theory in stochastic processes [58], σ_l is given by

$$\sigma_l = \frac{P_l \mathbb{E}[T]}{(\sum_{i=1}^M \alpha_i + \alpha_{\text{col}})\mathbb{E}[T] + \mathbb{E}[\mathbf{Idle}]} \tag{66}$$
$$= \frac{P_l \mathbb{E}[T]}{\mathbb{E}[T] + \frac{\mathbb{E}[T]}{\sum_{i=1}^M r_i}} \tag{67}$$
$$= \frac{[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^M r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{\sum_{i=1}^M r_i + 1}. \tag{68}$$

□

Appendix C
Proofs of Theorem 1, Corollary 2, Theorem 3, and Corollary 4.

C.1 The Proofs of Theorem 1 and Corollary 2

We prove Theorem 1 and Corollary 2 in three steps:

Step 1: We show that our solution $\mathbf{r}^\star$ (18) - (20) is feasible for Problem **1**.

Lemma 7. *If $\sum_{i=1}^M b_i \geq 1$, then the solution $\mathbf{r}^\star$ (18) - (20) is feasible for Problem **1**.*

Proof. See Appendix D. □

Hence, by substituting this solution $\mathbf{r}^\star$ into the objective function of Problem **1** in (17), we get an upper bound on the

optimal value $\bar{\Delta}_{\text{opt}}^{\text{w-peak}}$, which is expressed in the following lemma:

Lemma 8. *If* $\sum_{i=1}^{M} b_i \geq 1$, *then*

$$\bar{\Delta}_{opt}^{w\text{-}peak} \leq \bar{\Delta}^{w\text{-}peak}(\mathbf{r}^\star) \leq \sum_{i=1}^{M} \left[\frac{w_i e^{x^\star \frac{t_s}{\mathbb{E}[T]}} \left(1+\frac{1}{x^\star}\right)}{\min\{b_i, \beta^\star \sqrt{w_i}\}} + w_i \right], \tag{69}$$

where $x^\star$, $\beta^\star$ *are defined in* (19), (20).

Proof. In Lemma 7, we showed that our proposed solution $\mathbf{r}^\star$ (18) - (20) is feasible for Problem **1**. Hence, we substitute this solution into Problem **1** to obtain the following upper bound:

$$\sum_{i=1}^{M} \left[\frac{w_i e^{x^\star \frac{t_s}{\mathbb{E}[T]}} \left(1+\frac{1}{x^\star}\right) e^{-\min\{b_i, \beta^\star \sqrt{w_i}\} x^\star \frac{t_s}{\mathbb{E}[T]}}}{\min\{b_i, \beta^\star \sqrt{w_i}\}} + w_i \right]. \tag{70}$$

Next, we replace $e^{-\min\{b_i, \beta^\star \sqrt{w_i}\} x^\star (t_s/\mathbb{E}[T])}$ by 1 to derive another upper bound with a simple expression, which is given by (69). This completes the proof. □

Step 2: We now construct a lower bound on the optimal value of Problem **1**. Suppose that $\mathbf{r} = (r_1, \ldots, r_M)$ is a feasible solution to Problem **1**, such that $r_l > 0$ and

$$\frac{[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^{M} r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{\sum_{i=1}^{M} r_i + 1} \leq b_l, \forall l. \tag{71}$$

Because $[1 - e^{-r_l (t_s/\mathbb{E}[T])}] \sum_{i=1}^{M} r_i + r_l e^{-r_l (t_s/\mathbb{E}[T])} > r_l$ for all l, $\mathbf{r}$ satisfies $r_l / (\sum_{i=1}^{M} r_i + 1) \leq b_l$. Hence, the following Problem **2** has a larger feasible set than Problem **1**: (Problem **2**)

$$\bar{\Delta}_{\text{opt},2}^{\text{w-peak}} \triangleq \min_{r_l > 0} \sum_{l=1}^{M} \frac{w_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{r_l} e^{\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}} \left(1 + \sum_{i=1}^{M} r_i\right) + \sum_{l=1}^{M} w_l \tag{72}$$

$$\textbf{s.t. } r_l \leq b_l \left(\sum_{i=1}^{M} r_i + 1\right), \forall l, \tag{73}$$

where $\bar{\Delta}_{\text{opt},2}^{\text{w-peak}}$ is the optimal value of Problem **2**. The optimal objective value of Problem **2** is a lower bound of that of Problem **1**. We note that the constraint set corresponding to Problem **2** is convex. Thus, this relaxation converts the constraint set of Problem **1** to a convex one, and hence enables us to obtain a lower bound for the optimal value of Problem **1**, which is expressed in the following lemma:

Lemma 9. *If* $\sum_{i=1}^{M} b_i \geq 1$, *then*

$$\bar{\Delta}_{opt}^{w\text{-}peak} \geq \bar{\Delta}_{opt,2}^{w\text{-}peak} \geq \sum_{i=1}^{M} \left[\frac{w_i}{\min\{b_i, \beta^\star \sqrt{w_i}\}} + w_i \right], \tag{74}$$

where $\beta^\star$ *is the root of* (20).

Proof. See Appendix E. □

Step 3: After the upper and lower bounds of $\bar{\Delta}_{\text{opt}}^{\text{w-peak}}$ were derived in Steps 1-2, we are ready to analyze their gap. By combining (69) and (74), the sub-optimality gap of the solution $\mathbf{r}^\star$ (18) - (20) is upper bounded by

$$\left| \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) - \bar{\Delta}_{\text{opt}}^{\text{w-peak}} \right| \leq \sum_{i=1}^{M} \frac{w_i \left(e^{x^\star \frac{t_s}{\mathbb{E}[T]}} (1+\frac{1}{x^\star}) - 1 \right)}{\min\{b_i, \beta^\star \sqrt{w_i}\}}, \tag{75}$$

where $x^\star$, $\beta^\star$ are defined in (19), (20). Next, we characterize the right-hand-side (RHS) of (75) by Taylor expansion. For simplicity, let $\epsilon = \frac{t_s}{\mathbb{E}[T]}$. Using the expression for $x^\star$ from (19), we have

$$x^\star \epsilon = -\frac{\epsilon}{2} + \sqrt{\frac{\epsilon^2}{4} + \epsilon} = \frac{\epsilon}{\frac{\epsilon}{2} + \sqrt{\frac{\epsilon^2}{4} + \epsilon}} = \sqrt{\epsilon} + o(\sqrt{\epsilon}). \tag{76}$$

Moreover,

$$x^\star = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\epsilon}} = \frac{\frac{1}{\epsilon}}{\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\epsilon}}} = \frac{1}{\sqrt{\epsilon}} + o\left(\frac{1}{\sqrt{\epsilon}}\right). \tag{77}$$

Substituting (76) and (77) in (75), we obtain

$$\left| \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) - \bar{\Delta}_{\text{opt}}^{\text{w-peak}} \right| \tag{78}$$

$$\leq \sum_{i=1}^{M} \frac{w_i [e^{\sqrt{\epsilon} + o(\sqrt{\epsilon})} (1 + \sqrt{\epsilon} + o(\sqrt{\epsilon})) - 1]}{\min\{b_i, \beta^\star \sqrt{w_i}\}}$$

$$= \sum_{i=1}^{M} \frac{w_i [(1+\sqrt{\epsilon}+o(\sqrt{\epsilon}))(1+\sqrt{\epsilon}+o(\sqrt{\epsilon})) - 1]}{\min\{b_i, \beta^\star \sqrt{w_i}\}}$$

$$= 2\sqrt{\epsilon} \sum_{i=1}^{M} \frac{w_i}{\min\{b_i, \beta^\star \sqrt{w_i}\}} + o(\sqrt{\epsilon}), \tag{79}$$

where the second inequality involves the use of Taylor expansion. This proves Theorem 1.

We can observe that the gap $\left| \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) - \bar{\Delta}_{\text{opt}}^{\text{w-peak}} \right|$ in the energy-adequate regime converges to zero at a speed of $O(\sqrt{\epsilon})$, as $\epsilon \to 0$. Further, both the upper and lower bounds (69), (74), converge to $\sum_{i=1}^{M} [(w_i / \min\{b_i, \beta^\star \sqrt{w_i}\}) + w_i]$ as $t_s/\mathbb{E}[T] \to 0$. Thus, this value is the asymptotic optimal objective value of Problem **1**. This proves Corollary 2.

C.2 The Proofs of Theorem 3 and Corollary 4

Similar to Appendix C.1, we prove Theorem 3 and Corollary 4 also in three steps:

Step 1: We show that the proposed solution $\mathbf{r}^\star$ (18) and (25) - (27) is a feasible solution for Problem **1**.

Lemma 10. *If* $\sum_{i=1}^M b_i < 1$, *then the solution* $\mathbf{r}^\star$ (18) *and* (25) - (27) *is feasible for Problem* ***1****.*

Proof. See Appendix F. □

Now, we construct an upper bound on the optimal value of Problem **1** using our proposed solution as follows:

Lemma 11. *If* $\sum_{i=1}^M b_i < 1$, *then*

$$\bar{\Delta}_{opt}^{\text{w-peak}} \leq \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) \leq \sum_{l=1}^M \frac{w_l}{b_l} e^{\sum_{i=1}^M b_i x^\star \frac{t_s}{\mathbb{E}[T]}} \left(\frac{1}{x^\star} + \sum_{i=1}^M b_i\right) + \sum_{l=1}^M w_l, \tag{80}$$

where $x^\star$ *is defined in* (25).

Proof. In Lemma 10, we showed that our proposed solution $\mathbf{r}^\star$ (18) and (25) - (27) is feasible for Problem **1**. Hence, we substitute this solution into Problem **1** to obtain the following upper bound:

$$\sum_{l=1}^M \frac{w_l e^{-b_l x^\star \frac{t_s}{\mathbb{E}[T]}}}{b_l} e^{\sum_{i=1}^M b_i x^\star \frac{t_s}{\mathbb{E}[T]}} \left(\frac{1}{x^\star} + \sum_{i=1}^M b_i\right) + \sum_{l=1}^M w_l. \tag{81}$$

Next, we replace $e^{-b_l x^\star \frac{t_s}{\mathbb{E}[T]}}$ by 1 to derive another upper bound with a simple expression, which is given by (80). This completes the proof. □

Step 2: Similar to the proof in Appendix C.1, we use the relaxed problem, Problem **2**, to construct a lower bound as follows:

Lemma 12. *If* $\sum_{i=1}^M b_i < 1$, *then*

$$\bar{\Delta}_{opt}^{\text{w-peak}} \geq \bar{\Delta}_{opt,2}^{\text{w-peak}} \geq \sum_{l=1}^M \frac{w_l}{b_l} e^{\frac{-\sum_{i=1}^M b_i}{1-\sum_{i=1}^M b_i} \frac{t_s}{\mathbb{E}[T]}} + \sum_{l=1}^M w_l. \tag{82}$$

Proof. See Appendix G. □

Step 3: We now characterize the sub-optimality gap by analyzing the upper and lower bounds constructed above. By combining (80) and (82), the sub-optimality gap of the solution $\mathbf{r}^\star$ (18) and (25) - (27) is upper bounded by

$$\left|\bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) - \bar{\Delta}_{\text{opt}}^{\text{w-peak}}\right| \leq \sum_{l=1}^M \frac{w_l}{b_l} \left[e^{\sum_{i=1}^M b_i x^\star \frac{t_s}{\mathbb{E}[T]}} \left(\frac{1}{x^\star} + \sum_{i=1}^M b_i\right) - e^{\frac{-\sum_{i=1}^M b_i}{1-\sum_{i=1}^M b_i} \frac{t_s}{\mathbb{E}[T]}}\right]. \tag{83}$$

where $x^\star$ is defined in (25). Next, we characterize the RHS of (83) by Taylor expansion. For simplicity, let $\epsilon = t_s/\mathbb{E}[T]$, $Z = (\sum_{i=1}^M b_i)/(1-\sum_{i=1}^M b_i)$, and $k_l = (\sum_{i=1}^M b_i - b_l)/(1-\sum_{i=1}^M b_i)^2$. Using Taylor expansion, we are able to obtain the following:

$$\min_l c_l = 1 + \left(\min_l k_l\right)\epsilon + o(\epsilon), \tag{84}$$

$$\frac{1}{\min_l c_l} = \max_l \frac{1}{c_l} = 1 + \left(\max_l k_l\right)\epsilon + o(\epsilon). \tag{85}$$

Using (84), (85), $x^\star$ from (25), and Taylor expansion again, we get

$$\begin{aligned} e^{\sum_{i=1}^M b_i x^\star \epsilon} &= 1 + Z\left(1 + \left(\min_l k_l\right)\epsilon + o(\epsilon)\right)\epsilon + o(\epsilon) \\ &= 1 + Z\epsilon + o(\epsilon), \end{aligned} \tag{86}$$

$$\begin{aligned} \frac{1}{x^\star} + \sum_{i=1}^M b_i &= \frac{1-\sum_{i=1}^M b_i}{\min_l c_l} + \sum_{i=1}^M b_i \\ &= 1 + \left(\max_l k_l\right)\left(1 - \sum_{i=1}^M b_i\right)\epsilon + o(\epsilon), \end{aligned} \tag{87}$$

$$e^{-Z\epsilon} = 1 - Z\epsilon + o(\epsilon). \tag{88}$$

Substituting (86) - (88) into (83), we get (28). This proves Theorem (3).

Moreover, we observe that the gap $\left|\bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star) - \bar{\Delta}_{\text{opt}}^{\text{w-peak}}\right|$ in the energy-scarce regime converges to zero at a speed of $O(\epsilon)$, as $\epsilon \to 0$. Further, both the upper and lower bounds (80), (82), converge to $\sum_{i=1}^M [(w_i/b_i) + w_i]$ as $t_s/\mathbb{E}[T] \to 0$. Thus, this value is the asymptotic optimal objective value of Problem **1**. This proves Corollary 4.

APPENDIX D
PROOF OF LEMMA 7

First of all, we need to show that (20) has a solution for $\beta^\star$.

Lemma 13. *Suppose that* $w_l > 0$, *and* $b_l > 0$ *for all* l. *If* $\sum_{i=1}^M b_i \geq 1$, *then* (20) *has a unique solution on* $[0, \max_l(b_l/\sqrt{w_l})]$; *otherwise,* (20) *has no solution.*

Proof. It is clear that if $\sum_{i=1}^M b_i = 1$, then $\beta^\star$ satisfies (20) if and only if $\beta^\star \geq \max_l(b_l/\sqrt{w_l})$. Hence, (20) has a unique

solution on $[0, \max_l(b_l/\sqrt{w_l})]$ in this case. We now focus on the case of $\sum_{i=1}^{M} b_i > 1$. In this case, we have the following:

- If $\beta^\star = 0$, then $\sum_{i=1}^{M} \min\{b_i, \beta^\star\sqrt{w_i}\} = 0$.
- If $\beta^\star = \max_l(b_l/\sqrt{w_l})$, then $\sum_{i=1}^{M} \min\{b_i, \beta^\star\sqrt{w_i}\} > 1$.
- The left hand side (LHS) of (20) is strictly increasing and continuous in $\beta^\star$ on $[0, \max_l(b_l/\sqrt{w_l})]$.

As a result, (20) has a unique solution on $[0, \max_l(b_l/\sqrt{w_l})]$ in this case as well. Finally, if $\sum_{i=1}^{M} b_i < 1$, then $\sum_{i=1}^{M} \min\{b_i, \beta^\star\sqrt{w_i}\} \leq \sum_{i=1}^{M} b_i < 1$. Hence, (20) has no solution if $\sum_{i=1}^{M} b_i < 1$. This completes the proof. □

Since we have $\sum_{i=1}^{M} b_i \geq 1$, Lemma 13 implies that (20) has a solution for $\beta^\star$. Now, we are ready to prove Lemma 7. Consider the following constraints:

$$\frac{r_l \frac{t_s}{\mathbb{E}[T]} \sum_{i=1}^{M} r_i + r_l}{\sum_{i=1}^{M} r_i + 1} \leq b_l, \forall l. \tag{89}$$

Since we have

$$1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \leq r_l \frac{t_s}{\mathbb{E}[T]}, \tag{90}$$

$$e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \leq 1, \tag{91}$$

then,

$$[1 - e^{-r_l \frac{t_s}{\mathbb{E}[T]}}] \sum_{i=1}^{M} r_i + r_l e^{-r_l \frac{t_s}{\mathbb{E}[T]}} \leq r_l \frac{t_s}{\mathbb{E}[T]} \sum_{i=1}^{M} r_i + r_l. \tag{92}$$

Thus, if the constraints in (89) are satisfied for a given solution $\mathbf{r}$, then the constraints of Problem **1** are satisfied as well. We can observe that the constraints in (89) are equivalent to the following set of constraints:

$$\begin{aligned} r_l &\leq b_l \frac{x+1}{1 + \frac{t_s}{\mathbb{E}[T]} x}, \forall l \\ \sum_{i=1}^{M} r_i &= x. \end{aligned} \tag{93}$$

Now, it is easy to show that if $x \leq \sqrt{\mathbb{E}[T]/t_s}$, then $x \leq (x+1)/[1 + (t_s/\mathbb{E}[T])x]$. Meanwhile, our proposed solution $\mathbf{r}^\star$ (18) - (20) satisfies $\sum_{i=1}^{M} r_i^\star = x^\star$. Thus, if we can show that $x^\star \leq \sqrt{\mathbb{E}[T]/t_s}$, then

$$r_l^\star = \min\{b_l, \beta^\star\sqrt{w_l}\} x^\star \leq b_l x^\star \leq b_l \frac{x^\star + 1}{1 + \frac{t_s}{\mathbb{E}[T]} x^\star}, \tag{94}$$

and the constraints in (93) hold for our proposed solution $\mathbf{r}^\star$. What remains is to prove that $x^\star \leq \sqrt{\mathbb{E}[T]/t_s}$. We have

$$x^\star = \frac{-1}{2} + \sqrt{\frac{1}{4} + \frac{\mathbb{E}[T]}{t_s}} \tag{95}$$

$$= \frac{\frac{\mathbb{E}[T]}{t_s}}{\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\mathbb{E}[T]}{t_s}}} \tag{96}$$

$$\leq \frac{\frac{\mathbb{E}[T]}{t_s}}{\sqrt{\frac{\mathbb{E}[T]}{t_s}}} = \sqrt{\frac{\mathbb{E}[T]}{t_s}}. \tag{97}$$

Hence, our proposed solution $\mathbf{r}^\star$ (18) - (20) satisfies (93), which implies (89). This completes the proof. □

APPENDIX E
PROOF OF LEMMA 9

By replacing $e^{-r_l(t_s/\mathbb{E}[T])} e^{\sum_{i=1}^{M} r_i (t_s/\mathbb{E}[T])}$ in (72) of Problem **2** by 1, we obtain the following optimization problem:

$$\min_{r_l > 0} \sum_{l=1}^{M} \frac{w_l}{r_l} \left(1 + \sum_{i=1}^{M} r_i\right) + \sum_{l=1}^{M} w_l \tag{98}$$

$$\textbf{s.t. } r_l \leq b_l \left(\sum_{i=1}^{M} r_i + 1\right), \forall l. \tag{99}$$

Since $e^{-r_l(t_s/\mathbb{E}[T])} e^{\sum_{i=1}^{M} r_i (t_s/\mathbb{E}[T])} \geq 1$, Problem (98) serves as a lower bound of Problem **2**, and hence a lower bound of Problem **1** as well. Define an auxiliary variable $y = \sum_{i=1}^{M} r_i + 1$. By this, we solve a two-layer nested optimization problem. In the inner layer, we optimize $\mathbf{r}$ for a given y. After solving $\mathbf{r}$, we will optimize y in the outer layer. Now, fix the value of y, we obtain the following optimization problem (the inner layer):

$$\min_{r_i > 0} \sum_{i=1}^{M} \left[\frac{w_i y}{r_i} + w_i\right] \tag{100}$$

$$\textbf{s.t. } r_l \leq b_l y, \forall l, \tag{101}$$

$$\sum_{i=1}^{M} r_i + 1 = y. \tag{102}$$

The objective function in (100) is a convex function. Moreover, the constraints in (101) and (102) are affine. Hence, Problem (100) is convex. We use the Lagrangian duality approach to solve Problem (100). Problem (100) satisfies Slater's conditions. Thus, the Karush-Kuhn-Tucker (KKT) conditions are both necessary and sufficient for optimality [59]. Let $\gamma = (\gamma_1, \ldots, \gamma_M)$ and μ be the Lagrange multipliers associated with constraints (101) and (102), respectively.

Then, the Lagrangian of Problem (100) is given by

$$L(\mathbf{r},\gamma,\mu) = \sum_{i=1}^{M}\left[\frac{w_i y}{r_i}+w_i\right] + \sum_{i=1}^{M}\gamma_i(r_i - b_i y) + \mu\left(\sum_{i=1}^{M} r_i + 1 - y\right). \tag{103}$$

Take the derivative of (103) with respect to r_l and set it equal to 0, we get

$$\frac{-w_l y}{r_l^2} + \gamma_l + \mu = 0. \tag{104}$$

This and KKT conditions imply

$$r_l = \sqrt{\frac{w_l y}{\gamma_l + \mu}}, \tag{105}$$

$$\gamma_l \geq 0, r_l - b_l y \leq 0, \tag{106}$$

$$\gamma_l(r_l - b_l y) = 0, \tag{107}$$

$$\sum_{i=1}^{M} r_i + 1 = y. \tag{108}$$

If $\gamma_l = 0$, then $r_l = \sqrt{(w_l y)/\mu}$ and $r_l \leq b_l y$; otherwise, if $\gamma_l > 0$, then $r_l = b_l y$ and $r_l < \sqrt{(w_l y)/\mu}$. Hence, we have

$$r_l = \min\left\{b_l y, \sqrt{\frac{w_l y}{\mu^\star}}\right\}, \tag{109}$$

where by (102), $\mu^\star$ satisfies

$$\sum_{i=1}^{M}\min\left\{b_i y, \sqrt{\frac{w_i y}{\mu^\star}}\right\} + 1 = y. \tag{110}$$

We can observe that $\mu^\star$ is a function of y. Because of that, we can define $\beta^\star(y) = \sqrt{1/(y\mu^\star)}$, which is a function of y as well. Then, the optimum solution to (100) can be rewritten as

$$r_l = \min\{b_l, \beta^\star(y)\sqrt{w_l}\}y, \forall l, \tag{111}$$

where $\beta^\star(y)$ satisfies

$$\sum_{i=1}^{M}\min\{b_i, \beta^\star(y)\sqrt{w_i}\} + \frac{1}{y} = 1. \tag{112}$$

Substituting (111) and (112) back in Problem (100), we get the following optimization problem (the outer layer):

$$\min_{y>1}\sum_{i=1}^{M}\left[\frac{w_i}{\min\{b_i, \beta^\star(y)\sqrt{w_i}\}} + w_i\right] \tag{113}$$

$$\textbf{s.t.}\ \sum_{i=1}^{M}\min\{b_i, \beta^\star(y)\sqrt{w_i}\} + \frac{1}{y} = 1. \tag{114}$$

Problem (113) serves as a lower bound of Problem **2**, and hence a lower bound of Problem **1**. We can observe that the objective function in (113) is decreasing in $\beta^\star(y)$. Moreover, (114) implies that $\beta^\star(y)$ is strictly increasing in y if $\sum_{i=1}^{M} b_i \geq 1$. As a result, $y = \infty$ is the optimal solution of Problem (113). At the limit, the constraint (114) converges to (20). Since $\beta^\star$ serves as a solution for (20), we can deduce that $\lim_{y\to\infty}\beta^\star(y) = \beta^\star$. Thus, we have the following lower bound:

$$\bar{\Delta}_{\text{opt}}^{\text{w-peak}} \geq \bar{\Delta}_{\text{opt},2}^{\text{w-peak}} \geq \sum_{i=1}^{M}\left[\frac{w_i}{\min\{b_i, \beta^\star\sqrt{w_i}\}} + w_i\right]. \tag{115}$$

This completes the proof. □

APPENDIX F
PROOF OF LEMMA 10

Because $1 - e^{-x} \leq x$, we can obtain

$$\begin{aligned} & r_l e^{-r_l\frac{t_s}{\mathbb{E}[T]}} + [1 - e^{-r_l\frac{t_s}{\mathbb{E}[T]}}]\sum_{i=1}^{M} r_i \\ &= r_l + [1 - e^{-r_l\frac{t_s}{\mathbb{E}[T]}}]\left(\sum_{i=1}^{M} r_i - r_l\right) \\ &\leq r_l + r_l\frac{t_s}{\mathbb{E}[T]}\left(\sum_{i=1}^{m} r_i - r_l\right), \end{aligned} \tag{116}$$

Hence, if $\mathbf{r}$ satisfies the constraint

$$\frac{r_l + r_l\frac{t_s}{\mathbb{E}[T]}\left(\sum_{i=1}^{M} r_i - r_l\right)}{\sum_{i=1}^{M} r_i + 1} \leq b_l, \tag{117}$$

then $\mathbf{r}$ also satisfies the constraint of Problem **1** in (17). Consider the following set of solution indexed by a parameter $c > 0$:

$$r_l = cu_l,\ \forall l, \tag{118}$$

$$u_l = \frac{b_l}{1 - \sum_{i=1}^{M} b_i},\ \forall l \tag{119}$$

We want to find a c such that the solution in (118) and (119) is feasible for Problem **1**. To achieve this, we first substitute the solution (118) and (119) into the constraint (117), and get

$$\frac{cu_l + c^2 u_l\frac{t_s}{\mathbb{E}[T]}\left(\sum_{i=1}^{M} u_i - u_l\right)}{c\sum_{i=1}^{M} u_i + 1} \leq b_l. \tag{120}$$

If equality is satisfied in (120), we can obtain the following quadratic equation for c:

$$c^2\left[u_l\frac{t_s}{\mathbb{E}[T]}\left(\sum_{i=1}^{M} u_i - u_l\right)\right] + c\left(u_l - b_l\sum_{i=1}^{M} u_i\right) - b_l = 0. \tag{121}$$

The solution to (121) is given by c_l in (26). Hence, $r_l = c_l u_l$ is feasible for the constraint (117) for source l.

As feasibility for one source only is insufficient, we further prove that the solution in (118) and (119) with $c = \min_l c_l$ is feasible for satisfying the energy constraints of all sources $l = 1, \ldots, M$. To that end, let us consider the

monotonicity of the LHS of (120). By taking the derivative with respect to c, we get

$$\frac{u_l \frac{t_s}{\mathbb{E}[T]} \left(\sum_{i=1}^{M} u_i - u_l\right) \left(c^2 \sum_{i=1}^{M} u_i + 2c\right) + u_l}{(c \sum_{i=1}^{M} u_i + 1)^2} > 0. \quad (122)$$

Hence,

$$r_l = \left(\min_l c_l\right) u_l, \ \forall l, \quad (123)$$

is feasible for the energy constraints of all sources $l = 1, \ldots, M$. After some manipulations, the solution in (119) and (123) are equivalently expressed as (18) and (25) - (27). This completes the proof. □

APPENDIX G
PROOF OF LEMMA 12

By replacing $e^{-r_l(t_s/\mathbb{E}[T])}/r_l$ by $e^{-\sum_{i=1}^{M} r_i(t_s/\mathbb{E}[T])}/[b_l(\sum_{i=1}^{M} r_i + 1)]$ and $e^{\sum_{i=1}^{M} r_i(t_s/\mathbb{E}[T])}$ by 1 in (72) of Problem **2**, we obtain the following optimization problem:

$$\begin{aligned} \min_{r_l>0} & \sum_{l=1}^{M} \frac{w_l e^{-\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}}}{b_i} + \sum_{l=1}^{M} w_l \\ \textbf{s.t. } & r_l \leq b_l \left(\sum_{i=1}^{M} r_i + 1\right), \forall l. \end{aligned} \quad (124)$$

Since $r_l \leq b_l(\sum_{i=1}^{M} r_i + 1)$, we have

$$\frac{e^{-r_l \frac{t_s}{\mathbb{E}[T]}}}{r_l} \geq \frac{e^{-\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}}}{b_l \left(\sum_{i=1}^{M} r_i + 1\right)}. \quad (125)$$

Moreover, we have $e^{\sum_{i=1}^{M} r_i(t_s/\mathbb{E}[T])} \geq 1$. Thus, Problem (124) serves as a lower bound of Problem **2**, and hence a lower bound of Problem **1** as well. By removing the constant term $\sum_{l=1}^{M} w_l$ in the objective function of Problem (124) and then taking the logarithm, Problem (124) is reformulated as

$$\begin{aligned} \min_{r_i>0} & \log\left(\sum_{i=1}^{M} \frac{w_i}{b_i}\right) - \sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]} \\ \textbf{s.t. } & r_l \leq b_l \left(\sum_{i=1}^{M} r_i + 1\right), \forall l. \end{aligned} \quad (126)$$

Obviously, Problem (126) is a convex optimization problem and satisfies Slater's conditions. Thus, the KKT conditions are are necessary and sufficient for optimality. Let $\tau = (\tau_1, \ldots, \tau_M)$ be the Lagrange multipliers associated with the constraints of Problem (126). Then, the Lagrangian of Problem (126) is given by

$$\begin{aligned} L(\mathbf{r}, \tau) = & \log\left(\sum_{i=1}^{M} \frac{w_i}{b_i}\right) - \left(\sum_{i=1}^{M} r_i \frac{t_s}{\mathbb{E}[T]}\right) \\ & + \sum_{i=1}^{M} \tau_i \left[r_i - b_i \left(\sum_{i=1}^{M} r_i + 1\right)\right]. \end{aligned} \quad (127)$$

Take the derivative of (127) with respect to r_l and set it equal to 0, we get

$$\frac{-t_s}{\mathbb{E}[T]} + \tau_l(1 - b_l) - \sum_{i \neq l} \tau_i b_i = 0. \quad (128)$$

This and KKT conditions imply

$$\tau_l = \frac{t_s}{\mathbb{E}[T](1 - b_l)} + \frac{\sum_{i \neq l} \tau_i b_i}{1 - b_l}, \quad (129)$$

$$\tau_l \geq 0, r_l - b_l \left(\sum_{i=1}^{M} r_i + 1\right) \leq 0, \quad (130)$$

$$\tau_l \left[r_l - b_l \left(\sum_{i=1}^{M} r_i + 1\right)\right] = 0. \quad (131)$$

Since $\sum_{i=1}^{M} b_i < 1$, (129) implies that $\tau_l > 0$ for all l. This and (131) result in

$$r_l = b_l \left(\sum_{i=1}^{M} r_i + 1\right), \forall l. \quad (132)$$

Because $\sum_{i=1}^{M} b_i < 1$, (132) has a unique solution, which is given by

$$r_l = \frac{b_l}{1 - \sum_{i=1}^{M} b_i}, \forall l. \quad (133)$$

Hence, the solution to (124) and (126) is given by (133). Substitute (133) into (124), we get the following lower bound:

$$\bar{\Delta}_{\text{opt}}^{\text{w-peak}} \geq \bar{\Delta}_{\text{opt},2}^{\text{w-peak}} \geq \sum_{l=1}^{M} \frac{w_l e^{\frac{-\sum_{i=1}^{M} b_i}{1 - \sum_{i=1}^{M} b_i} \frac{t_s}{\mathbb{E}[T]}}}{b_l} + \sum_{l=1}^{M} w_l. \quad (134)$$

This completes the proof. □

APPENDIX H
PROOF OF COROLLARY 5

We start by solving Problem (38) for optimal $\mathbf{a}$. Problem (38) is a convex optimization problem and satisfies Slater's conditions. Thus, the KKT conditions are necessary and sufficient for optimality. Let $\lambda = (\lambda_1, \ldots, \lambda_M)$ and ν be the Lagrange multipliers associated with the constraints (39) and (40), respectively. Then, the Lagrangian of Problem (38)

is given by

$$L(\mathbf{a},\lambda,\nu)=\sum_{i=1}^{M}\left[\frac{w_i}{a_i}+w_i\right] + \sum_{i=1}^{M}\lambda_i(a_i-b_i)+\nu\left(\sum_{i=1}^{M}a_i-1\right). \tag{135}$$

Take the derivative of (135) with respect to a_l and set it equal to 0, we get

$$\frac{-w_l}{a_l^2}+\lambda_l+\nu=0. \tag{136}$$

This and KKT conditions imply

$$a_l=\sqrt{\frac{w_l}{\lambda_l+\nu}}, \tag{137}$$

$$\lambda_l\geq 0,\ a_l-b_l\leq 0, \tag{138}$$

$$\lambda_l(a_l-b_l)=0, \tag{139}$$

$$\nu\geq 0,\ \sum_{i=1}^{M}a_i-1\leq 0, \tag{140}$$

$$\nu\left(\sum_{i=1}^{M}a_i-1\right)=0. \tag{141}$$

If $\lambda_l=0$, then we have $a_l=\sqrt{w_l/\nu}$ and $a_l\leq b_l$. This implies that $\nu>0$ and hence $\sum_{i=1}^{M}a_i=1$, which holds when $\sum_{i=1}^{M}b_i\geq 1$.

If $\lambda_l>0$, then we have $a_l=b_l$ and $a_l\leq\sqrt{w_l/\nu}$. In this case, we either have $\nu>0$, which implies $\sum_{i=1}^{M}a_i=1$ and this holds when $\sum_{i=1}^{M}b_i\geq 1$; or $\nu=0$, which implies $\sum_{i=1}^{M}a_i\leq 1$ and this holds when $\sum_{i=1}^{M}b_i\leq 1$.

From the above argument, the solution can be driven according to the following two cases:

Case 1 (**Energy-adequate regime ($\sum_{i=1}^{M}b_i\geq 1$)**): In this case, the optimal solution is given by

$$a_l^\star=\min\left\{b_l,\sqrt{\frac{w_l}{\nu^\star}}\right\},\ \forall l, \tag{142}$$

where we must have $\nu^\star>0$, which implies $\sum_{i=1}^{M}a_i^\star=1$. Hence, $\nu^\star$ satisfies

$$\sum_{i=1}^{M}\min\left\{b_i,\sqrt{\frac{w_i}{\nu^\star}}\right\}=1. \tag{143}$$

By comparing (143) with (20), we can deduce that $\sqrt{1/\nu^\star}=\beta^\star$, where $\beta^\star$ satisfies

$$\sum_{i=1}^{M}\min\{b_i,\beta^\star\sqrt{w_i}\}=1. \tag{144}$$

Since $\sum_{i=1}^{M}b_i\geq 1$, (144) has a solution for $\beta^\star$ as shown in Lemma 13. Hence, the solution to Problem (38) can be rewritten as

$$a_l^\star=\min\{b_l,\beta^\star\sqrt{w_l}\},\ \forall l. \tag{145}$$

Substituting (145) into (38), we obtain

$$\bar{\Delta}_{\text{opt-s}}^{\text{w-peak}}=\sum_{i=1}^{M}\left[\frac{w_i}{\min\{b_i,\beta^\star\sqrt{w_i}\}}+w_i\right], \tag{146}$$

which is equal to the asymptotic optimal objective value of Problem **1** in energy-adequate regime in (24).

Case 2 (**Energy-scarce regime ($\sum_{i=1}^{M}b_i<1$)**): In this case, the optimal solution is

$$a_l^\star=b_l,\ \forall l. \tag{147}$$

Substituting by this into (38), we obtain

$$\bar{\Delta}_{\text{opt-s}}^{\text{w-peak}}=\sum_{i=1}^{M}\left[\frac{w_i}{b_i}+w_i\right], \tag{148}$$

which is equal to the asymptotic optimal objective value of Problem **1** in energy-scarce regime in (30). This completes the proof. □

APPENDIX I
PROOF OF THEOREM 6

I.1 Notation and Background on General State-Space Markov Processes

While analyzing learning algorithm, we will have to work with Markov processes on general state-space [55], [60]. In this section we provide a brief account of such processes.

Notation: For a set of r.v. s $\mathcal{X}$, we let $\mathcal{F}(\mathcal{X})$ denote the smallest sigma-algebra with respect to which each r.v. in $\mathcal{X}$ is measurable. For a set $\mathcal{X}$, we let $\mathcal{X}^c$ denote its complement. For an event $\mathcal{X}$, we let $\mathbb{1}(\mathcal{X})$ denote its indicator random variable. For a set $\mathcal{X}$, we let $\mathcal{B}(\mathcal{X})$ denote the sigma-algebra of Borel sets of $\mathcal{X}$.

We begin by showing that $\mathbf{s}(n)$ can be taken to be the system state /sufficient statistics [56] in order to describe the sampled process. In what follows, we let $\mathcal{S}:=\mathbb{R}_+\times\{0,1\}$. Denote by $\theta:=\int_{y=0}^{T_{\max}}y\,f(y)dy$ the mean transmission time of a packet of any source, i.e., we use the abbreviation $\theta=\mathbb{E}[T]$. The proof of the following result is omitted for brevity.

Lemma 14. *Consider the system in which M sources share a channel, and utilize the sleep period parameters as $\mathbf{r}(n)\equiv\mathbf{r}$ in order to modulate the sleep durations of sources. We then have that*

$$\mathbb{P}\left(\mathbf{s}(n+1)\in A\middle|\mathcal{F}_t\right)=\mathcal{K}(\mathbf{s}(n),\mathbf{r},A;f), \tag{149}$$

where $\mathcal{F}_t$ denotes the sigma-algebra generated by all the random variables until the n-th discrete sampling instant. The function

$\mathcal{K}$ is the kernel [55] associated with the controlled transition probabilities of the process $\mathbf{s}(n)$,

$$\mathcal{K} : \mathcal{S}^M \times \mathbb{R}_+^M \times \mathcal{B}(\mathbb{R}^M) \mapsto [0,1]. \tag{150}$$

Thus, $\mathcal{K}(\mathbf{s}, \mathbf{r}, A; f)$ is the probability with which the state at time $n+1$ belongs to the set A, given that the state at time n is equal to $\mathbf{s}$, and the vector comprising of sleep period parameters at time n is equal to $\mathbf{r}$. Note that the kernel is parametrized by the density function of transmission time f.

We begin by stating some definitions associated with Markov Chains on *General State-Spaces*. Though these can be found in standard textbooks on General State-Space Markov Chains such as [55], [60], we include them here in order to make the paper self-contained.

Let us now fix the controls at $\mathbf{r}(n) \equiv \mathbf{r}$, and consider the resulting discrete-time Markov chain $\mathbf{s}(n) \in \mathcal{S}^M$. If A is a Borel set, we let $P^n(\mathbf{x}, A)$ denote the probability of the event $\mathbf{s}(n) \in A$, given that $\mathbf{s}(0) = \mathbf{x}$.

Definition 1. *(Small Set) A set $C \in \mathcal{B}(\mathcal{S}^M)$ is called ν_m small if for all $\mathbf{x} \in C$ we have that*

$$P^m(\mathbf{x}, A) \geq \nu_m(A), \ \forall A \in \mathcal{B}(\mathcal{S}^M),$$

for some non-trivial measure $\nu_m(\cdot)$ and some $m \in \mathbb{N}$.

Definition 2. *(Petite Set) Let $\mathbf{q} = \{q_n\}_{n\in\mathbb{N}}$ be a probability distribution on $\mathbb{N}$. A set $C \in \mathcal{B}(\mathcal{S}^M)$ and a non-trivial sub-probability measure $\nu_q(\cdot)$ are called petite if we have that*

$$\sum_{n\in\mathbb{N}} q_n P^n(\mathbf{x}, A) \geq \nu_q(A), \forall A \in \mathcal{B}(\mathcal{S}^M), \forall x \in C.$$

Definition 3. *(Strong Aperiodicity) If there exists a ν_1 small set C such that we have $\nu_1(C) > 0$, then the chain $\mathbf{s}(n)$ is strongly aperiodic.*

I.2 Preliminary Results

We now show that in order to minimize the expected value of $C(H)$, it suffices to design controllers that "work directly" with the sampled system. Thus, the quantity $\mathbf{s}(n)$ as described in (42) serves as a sufficient statistic for the purpose of optimizing the expectation of cumulative peak age [56]. We also show that this objective can be posed as a constrained Markov decision process [61].

Lemma 15. *Let $\mathbf{s}(n), n = 1, 2, \ldots,$ be the sampled controlled Markov process. There exists a function $g : \mathcal{S}^M \mapsto \mathbb{R}$ so that $\mathbb{E}[C(H)]$ in (43) is given by $\mathbb{E}\left[\sum_{n=1}^{H} g(\mathbf{s}(n))\right]$.*

Proof. Consider the cumulative peak-age cost (43) in which the l-th source incurs a penalty of $\Delta_{l,i}^{\text{peak}}$ upon delivery of the i-th packet. Let this delivery occur at the end of the n-th discrete time-slot (note that this time n is random). Let us denote by $a_l^{\text{peak}}(n)$ the peak age of source l during the (continuous) time interval (in the non-discretized system) corresponding to the discrete time slots $n-1$ and n. We could (instead of charging a penalty of $\Delta_{l,i}^{\text{peak}}$ units at the end of n-th slot) charge the quantity $\mathbb{E}\left\{a^{peak}(n)|\mathbf{s}(n-1), \mathbf{r}(n-1)\right\}$ at the discrete time instant $n-1$. For sources $k \neq l$ that are not transmitting between $n-1$ and n, and have $m_k(n-1) = 0$, we let $g(\mathbf{s}(n-1)) = 0$. It then follows from the law of the iterated expectations [62] that the expected cost of the system under this modified cost function remains the same as that of the original system. This completes the proof. □

Ergodicity of $\mathbf{s}(n)$: We now derive a few useful results about the Markov process $\mathbf{s}(n)$.

Lemma 16. *Consider the multi-source wireless network operating under the controls $\mathbf{r}(n) \equiv \mathbf{r}$, and assume that the sensing time t_s is sufficiently small, i.e., it satisfies $t_s < 1$. Consider the associated process $\mathbf{s}(n), \ n = 1, 2, \ldots$ We then have the following:*

1) *Define*

$$e_i := (M-i)\epsilon, \text{ and } m_i = 0, \forall i \in [N],$$

where $\epsilon > 0$ is chosen to be sufficiently small. Consider the set

$$C := \otimes_{i=1}^{N} \left[[(M-i), (M-i)+e_i] \times \{m_i\}\right]. \tag{151}$$

The set C is small for the process $\mathbf{s}(n)$.

2) *For the process $\mathbf{s}(n)$, each compact set is petite.*
3) *The process $\mathbf{s}(n)$ is strongly aperiodic.*

Proof. 1) Consider $\mathbf{s}(0) \in C$. It follows that at time $n = 0$, all the sources are sleeping. Consider the following set denoted C': Sources 1 and 2 wake up within t_s time duration of each other, while the other sources wake up much later than these two. Consequently, there is a collision between Source 1 and Source 2, and hence at time $n = 1$ these two sources enter into sleep mode, so that at time $n = 1$ all the sources are asleep. Also assume that the cumulative time elapsed for this event to occur is approximately equal to $t_s + \delta$, where $\delta > 0$ is a sufficiently small parameter. The probability of the

event $\{\mathbf{s}(1) \in C'\}$ can be lower bounded as follows

$$\mathbb{P}(\mathbf{s}(1) \in C') \geq \left(r_1 \int_{\delta}^{\delta+\epsilon} \exp(-r_1 x)dx\right) \times t_s r_2 \exp(-r_2(\delta+\epsilon)) \times \left[\Pi_{i=3}^{N} \int_{\delta+\epsilon}^{\infty} r_i \exp(-r_i x)dx\right].$$

Since the above lower-bound on the probability of "reaching C'" is true for all $\mathbf{s}(0) \in C$, it follows from Definition 1 that the set C is small.

2) Consider the process $\mathbf{s}(n)$ starting in state $\mathbf{s}(0)$, and let the age vector $a(0)$ belong to a compact set, so that $\mathbf{s}(0)$ also belongs to a compact set. We will derive a lower bound on the probability of the event $\{\mathbf{s}(N) \in C\}$, where C is as in (151). This will prove (ii) since we have already shown in (i) that the set C is small. Consider the following sample path: at each time $i \in [1, M]$, we have that source i successfully transmits a packet, and moreover the age of the packet received is approximately equal to 1. We will derive a lower-bound on the probability of this event. In the following discussion we use $b > 0$ and $\eta \in (0, 1 - t_s - b)$, where η denotes the time when Source 1 wakes up. Since the counter of the i-th source has a probability density equal to $r_i e^{-r_i x}$, the probability that during the i-th slot source i gets channel access is lower bounded by $(1-\exp(-\eta r_i))\Pi_{j\neq i} e^{-r_j}$; while the probability that the age of its delivered packet is around 1, given that it wakes up at η, is lower bounded by $\int_0^b f(y)dy$. Thus, the probability of this sample path is lower bounded by

$$\Pi_{i=1}^{N}(1-\exp(-\eta r_i))\Pi_{j\neq i} e^{-r_j} \int_0^b f(y)dy.$$

This concludes the proof since along this sample path we have that $\mathbf{s}(N) \in C$.

3) It follows from the discussion on page 121 of [60] that in order to prove the claim it suffices to show that the volume of the set $C \cap C'$ is greater than 0. However, this condition holds true if the parameter δ in (i) above has been chosen so as to satisfy $t_s + \delta < \epsilon$.

□

We now show that the process $\mathbf{s}(n)$ has a certain "mixing property". For a measure μ and a function f, we define $\|\mu\|_f := \int f(x)d\mu(x)$.

Lemma 17 (Geometric Ergodicity)**.** *Consider the controlled Markov process $\mathbf{s}(n),\ n = 1, 2, \ldots$, associated with the network in which the controller utilizes $\mathbf{r}(n) \equiv \mathbf{r}$. The process $\mathbf{s}(n)$ has an invariant probability measure, which we denote as $\pi(\infty, \mathbf{r})$. Moreover,*

$$\int (\|\mathbf{y}\|_1 + 1)\, d\left(P^n(\mathbf{x}, \cdot) - \pi(\infty, \mathbf{r})\right)(\mathbf{y}) \leq R\left(\|\mathbf{s}(0)\|_1 + 1\right)\rho^n, n \in \mathbb{N}, \quad (152)$$

where $R > 0$, and $\rho < 1$.

Proof. Since we have shown in Lemma 16 that $\mathbf{s}(n)$ is strongly aperiodic, it follows from Theorem 6.3 of [60] that in order to prove the claim it suffices to show that the following holds true when $\|\mathbf{s}(n+1)\|_1$ is sufficiently large

$$\mathbb{E}\left(\|\mathbf{s}(n+1)\|_1 | \mathcal{F}_n\right) \leq \lambda \|\mathbf{s}(n)\|_1 + L, \quad (153)$$

where $\lambda < 1$. Note that each source gets to transmit with a probability at least $\min_l \alpha_l$, and also the expected value of the inter-sampling time is upper-bounded by $\max\left\{\mathbb{E}[T], \frac{\mathbb{E}[T]}{\sum_{i=1}^{M} r_i} + t_s\right\}$. It then follows that (153) holds true with λ set equal to $\min_l \alpha_l$, and L equal to $\max\left\{\mathbb{E}[T], \frac{\mathbb{E}[T]}{\sum_{i=1}^{M} r_i} + t_s\right\}$. □

Lemma 18. *(Differential Cost Function) Consider the process $\mathbf{s}(n),\ n = 1, 2, \ldots$, that describes the evolution of the network in which the controller utilizes $\mathbf{r}(n) \equiv \mathbf{r}$. Then, there exists a function $V : \mathcal{S}^M \mapsto \mathbb{R}$ that satisfies*

$$V(\mathbf{x}) + \int g(\mathbf{x}) d\pi(\infty, \mathbf{r}) = g(\mathbf{x}) + \int \mathcal{K}(\mathbf{x}, \mathbf{r}, y; f) V(y) dy, \quad (154)$$

where $\mathcal{K}$ is the transition kernel as described in Lemma 14, the function g is the one-step cost function as in Lemma (15)*. Moreover, the function V satisfies the following,*

$$V(\mathbf{x}) \leq \frac{R}{1-\rho}\left(\|\Delta(0)\|_1 + 1\right), \quad (155)$$

where the constant R is as in Lemma 17.

Proof. We have shown in Lemma 17 that the process $\mathbf{s}(n)$ is geometrically ergodic. Hence, it follows from Theorem 7.5.10 of [63] that there exists a function $V(\cdot)$ that satisfies (154), and moreover it is given as follows,

$$V(\mathbf{x}) = \sum_{n=1}^{\infty} \left[\mathbb{E}_x\left(g(\mathbf{x}(n))\right) - \int_{\mathcal{S}^M} g(\mathbf{y}) d\pi(\infty, \mathbf{r})(\mathbf{y})\right], x \in \mathcal{S}.$$

Substituting the geometric bound (152) into the above, we

obtain the following

$$\begin{aligned} V(\mathbf{x}) &= \sum_{n=1}^{\infty} \left[\mathbb{E}_{\mathbf{x}}\, g(\mathbf{x}(n)) - \int_{\mathcal{S}^M} g(\mathbf{y}) d\pi(\infty, \mathbf{r})(\mathbf{y}) \right] \\ &\leq \sum_{n=1}^{\infty} \left| \mathbb{E}_{\mathbf{x}}\, g(\mathbf{x}(n)) - \int_{\mathcal{S}^M} g(\mathbf{y}) d\pi(\infty, \mathbf{r})(\mathbf{y}) \right| \\ &\leq R\left(\|\mathbf{x}(0)\|_1 + 1\right) \sum_{n=1}^{\infty} \rho^n \\ &= \frac{R\left(\|\mathbf{x}(0)\|_1 + 1\right)}{1-\rho}, \end{aligned}$$

where $\rho < 1$. □

Lemma 19 (Smoothness properties of the optimal average cost)**.** *The optimal sleep period parameters $\mathbf{r}_\theta^\star$ and average cost $\bar{\Delta}^{w-peak}$ satisfy the following:*

1) *We have that the function $\mathbf{r}_\theta^\star : \Theta \mapsto \mathbb{R}_+^M$ that maps the mean transmission time θ to the optimal sleep period parameter, is a continuous function of θ. Similarly, the average peak age is a continuous function of $\mathbf{r}$, i.e.,*

$$\begin{aligned} &\lim_{\mathbf{r}\to\mathbf{r}_\theta^\star} \lim_{H\to\infty} \frac{1}{H} \sum_{n=1}^{H} \mathbb{E}_{\mathbf{r}} \left[g(\mathbf{s}(n)) \right] \\ &\to \lim_{H\to\infty} \frac{1}{H} \sum_{n=1}^{H} \mathbb{E}_{\mathbf{r}_\theta^\star} \left[g(\mathbf{s}(n)) \right], \end{aligned}$$

   where the sub-script $\mathbf{r}$ in the expectation $\mathbb{E}_{\mathbf{r}}$ above refers to the fact that the averaging is performed w.r.t. the measure induced by the policy that uses sleep rates equal to $\mathbf{r}$.
2) *The cumulative peak-age is locally Lipschitz continuous function of $\mathbf{r}$. Thus,*

$$|\bar{\Delta}^{w-peak}(\mathbf{r}_\theta^\star; \theta) - \bar{\Delta}^{w-peak}(\mathbf{r}; \theta)| \leq L_1 \|\mathbf{r}_\theta^\star - \mathbf{r}\|,$$

   whenever $\|\mathbf{r}_\theta^\star - \mathbf{r}\|$ is sufficiently small, and where the Lipschitz constant at sleep period parameter $\mathbf{r}$ is given by

$$L_1 := \max_{i\in[M]} \frac{\partial \bar{\Delta}^{w-peak}}{\partial r_i}(\mathbf{r}).$$

   Similarly, the optimal sleep period parameter is a locally Lipschitz function of θ, so that we have,

$$\|\mathbf{r}_{\theta_1}^\star - \mathbf{r}_{\theta_2}^\star\| \leq L_2 |\theta_1 - \theta_2|, L_2 > 0,$$

   whenever $|\theta_1 - \theta_2|$ is sufficiently small.

In summary, there exists a $\delta > 0$ such that whenever $|\theta_1 - \theta_2| \leq \delta$, then

$$|\bar{\Delta}^{w-peak}(\mathbf{r}_{\theta_1}^\star; \theta) - \bar{\Delta}^{w-peak}(\mathbf{r}_{\theta_2}^\star; \theta)| \leq L |\theta_1 - \theta_2|.$$

Proof. Continuity of the functions under discussion is immediate from the relations (18), (19), (20), (25), (26), (27). To prove the statement about Lipschitz continuity, it suffices to show that the average peak age is a Lipschitz continuous function of $\mathbf{r}$, and the optimal rate $\mathbf{r}_\theta^\star$ is Lipschitz continuous function of θ. To prove this, it suffices to show that the average peak age is a continuously differentiable function of $\mathbf{r}$, and also $\mathbf{r}_\theta^\star$ is a continuously differentiable function of θ (see [64] for more details). The continuously differentiable property is evident from the relations (11), (18)-(20) and (25)-(27). This completes the proof. □

Bounds on the Estimation Error: We now derive some concentration results for the estimate $\hat{\theta}(n)$ around the true value $\theta^\star$. Let $\mathcal{C}(n)$ be the confidence interval associated with the estimate $\hat{\theta}(n)$, i.e.,

$$\mathcal{C}(n) := \left\{ \theta : |\theta - \hat{\theta}(n)| \leq \xi(n), \theta > 0 \right\}, \tag{156}$$

where

$$\xi(n) := T_{\max} \sqrt{\frac{2\log\left(n^\gamma\right)}{N(n)}}, 1 \leq n \leq H,$$

$\gamma \geq 4$ is a constant, $N(n)$ is the total number of packet deliveries until n, and $T_{\max}$ is the maximum possible transmission time. We begin by showing that with a high probability, our confidence balls are true at all the times.

Lemma 20. *Define*

$$\mathcal{G}_1(n) := \left\{ \omega : \theta^\star \in \mathcal{C}(n) \right\},$$

where $\mathcal{C}(n)$ is as in (156)*, and $\theta^\star$ is the vector consisting of true parameter values. We then have that*

$$\mathbb{P}\left(\mathcal{G}_1^c(n)\right) \leq \frac{1}{n^{\gamma-1}}.$$

Proof. Fix a positive integer n_0, and let $\hat{\theta}$ denote the empirical estimate obtained from n_0 samples $T(1), T(2), \ldots, T(n_0)$ of the service times. It follows from Azuma-Hoeffding's inequality [65] that

$$\mathbb{P}\left(|\hat{\theta} - \theta^\star| > x \right) \leq \exp\left(-\frac{n_0 x^2}{2T_{\max}^2} \right).$$

By using $x = T_{\max}\sqrt{\frac{2\log(n^\gamma)}{n_0}}$ in the above, we obtain,

$$\begin{aligned} \mathbb{P}\left(|\hat{\theta} - \theta^\star| > T_{\max}\sqrt{\frac{\log n^\gamma}{n_0}} \right) &\leq \exp\left(-\log n^\gamma\right) \\ &= \frac{1}{n^\gamma}. \end{aligned}$$

Since the total number of samples n_0 can assume values from the set $\{0, 1, 2, \ldots, n\}$, the proof then follows by using union bound on n_0. □

Lemma 21. *Fix a $\delta_1 \in (0, p_{\min})$, where $p_{\min}$ is as in* (51)*. Define*

the event,

$$\mathcal{G}_2(n) := \left\{\omega : N(n) > (p_{\min} - \sqrt{\delta_1})n\right\}, \tag{157}$$

where $N(n)$ denotes the number of samples that have been obtained until time n for estimating transmission times. We then have that

$$\mathbb{P}(\mathcal{G}_2^c(n)) \leq \exp\left(-\delta_1 n\right).$$

Proof. Consider the following martingale difference sequence $m(i) = \mathbb{E}\left\{c(i)\middle|\mathcal{F}_{i-1}\right\} - c(n)$. Since $\mathbb{E}\left\{c(i)\middle|\mathcal{F}_{i-1}\right\} \geq p_{\min}$, we have that

$$\sum_{i=1}^{n} m(i) \geq c_{\min} n - N(n). \tag{158}$$

Since $|m(i)| \leq 1$, we have the following from Azuma-Hoeffding's inequality [65],

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} m(i)\right| \geq x\right) \leq \exp\left(-\frac{x^2}{n}\right).$$

Letting $x = \sqrt{\delta_1} n$, we get the following,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} m(i)\right| \geq \sqrt{\delta_1} n\right) \leq \exp\left(-\delta_1 n\right). \tag{159}$$

Substituting (158) into the above inequality, we obtain

$$\mathbb{P}\left(N(n) \leq \left(p_{\min} - \sqrt{\delta_1}\right) n\right) \leq \exp\left(-\delta_1 n\right).$$

This completes the proof. □

I.3 Regret Analysis

The cumulative regret $R(H)$ (45) decomposes into the sum of "episodic regrets" $R^{(e)}(k)$ as follows:

$$\mathbb{E}\left[R(H)\right] = \sum_{k=1}^{K} \mathbb{E}\left[\, R^{(e)}(k)\right], \tag{160}$$

$$\text{where } R^{(e)}(k) := \mathbb{E}\left\{\sum_{n\in\mathcal{E}_k} g(\mathbf{s}(n)) - \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star)\middle|\mathcal{F}_{\tau_k}\right\}. \tag{161}$$

Combining the regret decomposition with the smoothness properties of the optimal average cost that were derived in Lemma 19, we obtain the following key result that allows us to upper-bound $R(H)$.

Lemma 22. *The cumulative expected regret* (160) *for a learning algorithm can be upper-bounded as follows,*

$$\mathbb{E}\left[R(H)\right] \leq K_2 \sum_{k=1}^{K} (\tau_{k+1} - \tau_k)\, \mathbb{P}(|\hat{\theta}(\tau_k) - \theta^\star| > \delta) + L \sum_{k=1}^{N} (\tau_{k+1} - \tau_k) \mathbb{E}\left(|\hat{\theta}(\tau_k) - \theta^\star| \mathbb{1}\left\{|\hat{\theta}(\tau_k) - \theta^\star| \leq \delta\right\}\right), \tag{162}$$

where the constant $\delta > 0$ is as in Lemma 19.

Proof. It follows from the ergodicity properties of the process $\mathbf{s}(n)$ that were proved in Lemma 18 and Assumption 2 regarding $\mathbf{s}(n)$, that the episodic regret can be bounded as follows (ρ, R are as in Lemma 18 and Assumption 2),

$$R^{(e)}(k) \leq \frac{R}{1-\rho}(K_1 + 1) \tag{163}$$

$$+ \left|\bar{\Delta}^{w-peak}(\mathbf{r}^\star_{\hat{\theta}(\tau_k)}; \theta) - \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star)\right| (\tau_{k+1} - \tau_k). \tag{164}$$

The following two events are possible:

(i) $|\theta^\star - \hat{\theta}(\tau_k)| < \delta$: In this case it follows from Lemma 19 that

$$\left|\bar{\Delta}^{w-peak}(\mathbf{r}^\star_{\hat{\theta}(\tau_k)}; \theta) - \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star)\right| \leq L|\theta^\star - \hat{\theta}(\tau_k)|.$$

(ii) $|\theta^\star - \hat{\theta}(\tau_k)| > \delta$: It follows from Assumption 3 that the average performance under any sleep parameter cannot exceed K_2, and hence we can bound $\left|\bar{\Delta}^{w-peak}(\mathbf{r}^\star_{\hat{\theta}(\tau_k)}; \theta) - \bar{\Delta}^{\text{w-peak}}(\mathbf{r}^\star)\right|$ by K_2.

The proof then follows by substituting the bounds discussed above for the two cases into (163), and using regret decomposition result. □

We now separately bound the expressions obtained in the two events ($|\theta^\star - \hat{\theta}(\tau_k)| < \delta$, $|\theta^\star - \hat{\theta}(\tau_k)| > \delta$).

Regret when $|\theta^\star - \hat{\theta}(\tau_k)| > \delta$:

Choose a sufficiently large $k_0 \in \mathbb{N}$ that satisfies

$$\tau_{k_0} = O\left(\frac{1}{\delta_1} \log H\right). \tag{165}$$

Define the following event

$$\mathcal{G}_3 := \cap_{k \geq k_0} \mathcal{G}_2(\tau_k).$$

By combining the result of Lemma 21 with the union bound and using (165) we conclude that $\mathcal{G}_3$ has a probability greater than $1 - \sum_{k>k_0} \exp(-\delta_1 \tau_k) = 1 - O\left(\frac{1}{H}\right)$. On $\mathcal{G}_3$, the number of samples $N(\tau_k)$ at the beginning of each episode $k > k_0$ is greater than $\left(p_{\min} - \sqrt{\delta_1}\right) \tau_k$. Thus on

$\mathcal{G}_3$, for episodes $k > k_0$ the radius of $\mathcal{C}(\tau_k)$ is less than $\sqrt{\frac{\gamma \log H}{(p_{\min}-\sqrt{\delta_1})\tau_k}}$. Let k_1 be the smallest integer that satisfies

$$\frac{\gamma \log H}{(p_{\min}-\sqrt{\delta_1})\tau_{k_1}} \leq \delta^2, \text{ i.e. } \tau_{k_1} \geq \frac{1}{(p_{\min}-\sqrt{\delta_1})\delta^2}\gamma \log H, \tag{166}$$

where the constant $\delta > 0$ is as in Lemma 19. Thus on $\mathcal{G}_3$, for episodes $k \geq \max\{k_0, k_1\}$, the radius of confidence intervals is less than δ. Note that on $\cap_k \mathcal{G}_1(\tau_k)$ the confidence intervals (156) at the beginning of each episode are true. Hence, on $\{\cap_k \mathcal{G}_1(\tau_k)\} \cap \mathcal{G}_3$ we have $|\hat{\theta}(\tau_k) - \theta^\star| < \delta$ for epsiodes $k \geq \max\{k_0, k_1\}$. Thus, on $\{\cap_k \mathcal{G}_1(\tau_k)\} \cap \mathcal{G}_3$ this regret is bounded by $K_2 \max\{\tau_{k_0}, \tau_{k_1}\}$. Now consider sample paths for which some of the confidence intervals fail. The probability that $\mathcal{C}(\tau_k)$ fails is less than $\frac{1}{\tau_k^{\gamma-1}}$ (Lemma 20); moreover since the episode duration of $\mathcal{E}_k$, $(\tau_{k+1}-\tau_k)$ is less than τ_k, we have that the expected value of the regret during $\mathcal{E}_k$ in the event of failure of $\mathcal{C}(\tau_k)$ is less than $K_2 \frac{1}{\tau_k^{\gamma-2}}$. Since $\gamma \geq 4$, the cumulative expected regret arising from this is bounded by $K_2 \sum_k \frac{1}{\tau_k^{\gamma-2}} \leq K_2 \frac{\pi^2}{6}$ [66]. We summarize our discussion as follows.

Lemma 23. *Under Algorithm 2 the following is true,*

$$\sum_{k=1}^{K} (\tau_{k+1}-\tau_k)\, \mathbb{P}(|\hat{\theta}(\tau_k) - \theta^\star| > \delta) \leq K_2 \max\left\{ \frac{\gamma \log H}{(p_{\min}-\sqrt{\delta_1})\delta^2}, O\left(\frac{1}{\delta_1}\log H\right)\right\} + K_2 \frac{\pi^2}{6}, \tag{167}$$

where $\gamma \geq 4$.

Regret when $|\theta^\star - \hat{\theta}(\tau_k)| < \delta$:

As discussed above, on $\cap_k \mathcal{G}_1(\tau_k) \cap \mathcal{G}_3$ we have $|\theta(\tau_k) - \theta^\star| < \delta$ for episodes $k > k_1$. Thus, after using the smoothness property of optimal average cost that was developed in Lemma 19, we obtain that the second summation in the r.h.s. of (162) can be bounded by the following quantity,

$$\sum_{k>k_1} (\tau_{k+1}-\tau_k) \sqrt{\frac{\gamma \log H}{(p_{\min}-\sqrt{\delta_1})\tau_k}}.$$

Since we have $\tau_{k+1} - \tau_k \leq \tau_k$, the above can be bounded by $\sqrt{\frac{\gamma \log H}{(p_{\min}-\sqrt{\delta_1})}} \sum_{k>k_1} \sqrt{\tau_k}$. By using Cauchy Schwart'z inequality, the quantity $\sum_{k>k_1} \sqrt{\tau_k}$ can be upper-bounded as $\sqrt{HK}$, where K denotes the number of episodes. Since $K = O(\log H)$, this regret is bounded by $\sqrt{\frac{H\gamma(\log H)^2}{(p_{\min}-\sqrt{\delta_1})}}$. The bound we discussed is summarized below.

Lemma 24. *Under Algorithm 2 the following is true,*

$$L \sum_{k=1}^{N} (\tau_{k+1}-\tau_k) \mathbb{E}\left(|\hat{\theta}(\tau_k) - \theta^\star| \mathbb{1}\left\{ |\hat{\theta}(\tau_k) - \theta^\star| \leq \delta \right\}\right) \leq L \sqrt{\frac{H\gamma(\log H)^2}{(p_{\min}-\sqrt{\delta_1})}}. \tag{168}$$

We are now in a position to prove main result Theorem 6.

Proof. (Theorem 6) The proof follows by substituting the bounds obtained in Lemma 23 and Lemma 24 into the regret decomposition result of Lemma 22. □