

# Can Decentralized Status Update Achieve Universally Near-Optimal Age-of-Information in Wireless Multiaccess Channels?

Zhiyuan Jiang<sup>1</sup>, Bhaskar Krishnamachari<sup>2</sup>, Sheng Zhou<sup>1</sup>, Zhisheng Niu<sup>1</sup>,  
<sup>1</sup>{zhiyuan, sheng.zhou, niuzhs}@tsinghua.edu.cn, Tsinghua University, Beijing, China  
<sup>2</sup> bkrishna@usc.edu, University of Southern California, Los Angeles, USA

**Abstract**—In an Internet-of-Things system where status data are collected from sensors and actuators for time-critical applications, the freshness of data is vital and can be quantified by the recently proposed age-of-information (AoI) metric. In this paper, we first consider a general scenario where multiple terminals share a common channel to transmit or receive randomly generated status packets. The optimal scheduling problem to minimize AoI is formulated as a restless multi-armed bandit problem. To solve the problem efficiently, we derive the Whittle's index in closed-form and establish the indexability thereof. Compared with existing work, we extend the index policy for AoI optimization to incorporate stochastic packet arrivals and optimal packet management (buffering the latest packet). Inspired by the index policy which has near-optimal performance but is centralized by nature, a decentralized status update scheme, i.e., the index-prioritized random access policy (IPRA), is further proposed, achieving universally near-optimal AoI performance and outperforming state-of-the-arts in the literature.

**Index Terms**—Internet-of-Things, age-of-information, Markov decision process, restless multi-armed bandit, Whittle's index, random access

## I. INTRODUCTION

The freshness, or timeliness, of information maintained at interested nodes is critical in many real-world applications, such as a central controller which requires time-sensitive status parameters from sensors, a cellular base station which utilizes channel state information for efficient transmissions in fast-fading scenarios, and distributed actuators where each actuator requires independent and timely inputs. With the emergence of Internet-of-Things (IoT), such information freshness is becoming increasingly important; a reasonable quantification thereof is the recently proposed age-of-information (AoI) [1]–[5]. Simply put, it denotes the difference between the current time and the status (maintained currently) generation time, i.e.,

$$h(t) \triangleq t - \mu(t), \quad (1)$$

where  $h(t)$  denotes the AoI at time  $t$  and  $\mu(t)$  denotes the generation time of the status maintained at destination at time  $t$ . In scenarios where a more up-to-date status renders an old status useless, i.e., the system status evolution is Markovian such that the current status is irrelevant with any old status given a newer status, AoI provides a compressive, and still mathematical tractable, characterization of the information freshness.

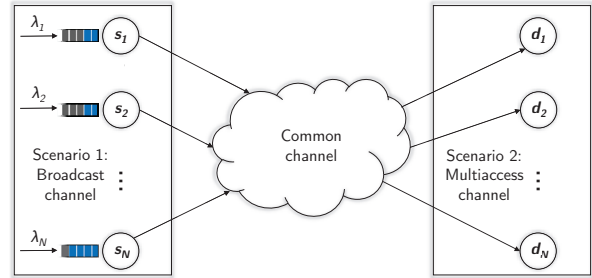


Fig. 1. A general status update system architecture with  $N$  source-destination pairs sharing one common channel.

There have been extensive efforts in the literature for AoI optimization in various scenarios. Considering a single source-destination pair, the queuing theory has been applied by many works to analyze and optimize the AoI performance (cf. [1], [2], [4]); the problem of scheduling multiple source-destination pairs has been mainly formulated and addressed with Markov decision process (MDP) and restless multi-armed bandit (RMAB) under different channel models and system assumptions [6]–[8]; a considerable amount of efforts have also been dedicated to, e.g., considering energy harvesting sources, general service processes ( $M/G/1$ ) and so forth (cf. a compressive and timely paper collection on [9]).

In this paper, firstly towards obtaining the optimal performance, we consider a general status update scheduling problem where multiple source-destination pairs with randomly generated status packets share a common channel which is depicted in Fig. 1; note that both wireless broadcast channels and multiaccess channels can be incorporated in this model by considering co-located source nodes and destination nodes, respectively. Our goal in this part is to find the optimal (centralized) status update policy, including the scheduling decision and packet management policy, which minimizes the infinite-horizon time-average AoI. Towards this end, we derive the Whittle's index [10] of the problem in closed-form and establish the corresponding indexability. Based on Weber and Weiss [11] and numerous evaluation results in the literature, the Whittle's index policy is considered near-optimal, especially when the number of terminals is large. Furthermore, we then focus on the multiple-access channel

which is of key interests in future IoT systems. A decentralized status update scheme, i.e., the index-prioritized random access policy (IPRA), is proposed based on the derived Whittle's index and shows universally near-optimal performance.

### A. Related Work

The decentralized scheduling problem for AoI optimization is considered by Jiang *et al.* [8], where a round-robin policy with one-packet buffers policy (RR-ONE) is shown to achieve asymptotically optimal performance with closed-form achievable AoI expressions. However, in the non-asymptotic regime, RR-ONE exhibits notable performance degradation whereas our proposed IPRA is universally near-optimal.

The Whittle's index has been adopted to solve the *centralized* scheduling problem for AoI optimization [6], [7]. Kadota *et al.* [6] address the deterministic packet arrival scenario comprehensively where they derive the index policy and also prove its performance bound. We adopt similar methodologies and generalize to stochastic packet arrival scenarios. In parallel with this work, we find that Hsu [7] also derives the Whittle's index with random packet arrivals; however, a specific no-buffer packet management policy is considered whereby the age of a packet at terminals either does not exist (no arrival at current time slot) or equals one (a packet just arrived). Our work generalizes the work [7] to allow arbitrary packet buffering policy and shows evident performance gain by using only one-packet buffers. Moreover, the index derivation in this paper is more challenging due to the fact that a two-dimensional system state is involved (the states in [7] are considered "nearly" one-dimensional since the age of packets is either zero or one). Additionally, it is shown that the index expression in [7, Theorem 7] coincides with a special case of our results.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a general scenario where one central controller is either collecting or broadcasting status update packets for multiple terminals. There are a total of  $N$  terminals, e.g., sensors and actuators. A time slotted system is considered. The status updates are conveyed by randomly generated packets at each terminal, reflecting the current status information sensed by terminals and stored at terminal queues. The packet arrivals are modeled by independently identically distributed (i.i.d.) Bernoulli processes with mean rates  $\lambda_n \in [0, 1]$ ,  $\forall n = \{1, \dots, N\}$ ; unlike conventional systems, these packets are usually short, containing a small amount of information however requiring very stringent timeliness. Therefore, instead of throughput, we adopt the AoI metric as our main optimization target. Concretely, the  $\tau$ -horizon time-average AoI of the system is defined by

$$\Delta_{\pi}^{(\tau, N)} \triangleq \frac{1}{\tau N} \sum_{t=1}^{\tau} \sum_{n=1}^N \mathbb{E}[h_{n, \pi}(t)], \quad (2)$$

where  $\pi$  denotes an admissible policy,  $\tau$  is the time horizon length, and  $h_{n, \pi}(t)$  denotes the AoI of terminal- $n$  at the  $t$ -th

time slot under policy  $\pi$ . In particular, the long-time-average AoI of the system is concerned, which is defined by

$$\bar{\Delta}_{\pi, N} \triangleq \limsup_{\tau \rightarrow \infty} \Delta_{\pi}^{(\tau, N)}. \quad (3)$$

### A. Status Update Process

In order to minimize their AoI, the terminals should decide on a transmission scheduling scheme by which they can update the status in a timely fashion and also avoid collisions. In principle, only one terminal should be scheduled for collision avoidance; however, considering the fact that decisions are made autonomously and terminals cannot coordinate perfectly based on global state information, we include concurrent transmissions in our model and hence possible collisions. The status update decisions include:

- *Scheduled terminal set and transmission probabilities:* Decide the set of terminals that is scheduled and the transmission probability  $p_i$  of each terminal.
- *Packet management:* Once a terminal is scheduled, a status update packet is then transmitted based on a packet management policy; the terminal can choose which packet in its buffer to transmit.

Due to the decoupled packet management and terminal scheduling, it is obvious that the optimal packet management scheme is to transmit the most up-to-date packet; this is equivalent to maintaining a one-packet buffer at each terminal and only retains the newest packet. Note that the one-buffer packet management is not necessarily optimal when considering service interruption which, however, does not exist in this work [12].

The evolution of the AoI of terminal- $n$  can be written as

$$h_{n, \pi}(t+1) = h_{n, \pi}(t) + 1 - u_{n, \pi}(t) \prod_{j \neq n} (1 - u_{j, \pi}(t)) g_{n, \pi}(t), \quad (4)$$

where  $u_{n, \pi}(t) = 1$  denotes the terminal- $n$  transmits in this time slot and zero otherwise, and the AoI reduction is denoted by  $g_{n, \pi}(t)$  which equals the time duration (time slots) between the generation of the last received packet from terminal- $n$  and the updated packet's generation time, i.e.,

$$g_{n, \pi}(t) = h_{n, \pi}(t) - a_{n, \pi}(t), \quad (5)$$

where  $a_{n, \pi}(t)$  denotes the age of the packet at terminal- $n$  in the  $t$ -th time slot. Note that  $g_{n, \pi}(t)$  equals zero if terminal- $n$  has no packet to update; to unify the notation, we prescribe  $a_{n, \pi}(t) = h_{n, \pi}(t)$  in this circumstance and hence  $g_{n, \pi}(t) \geq 0$ .

To be clear, the sequence of events is illustrated in Fig. 2. The AoI (the cost in the MDP formulation introduced later) at

Scheduling decision	Packet transmission and status update	New packet arrival	Age $\leftarrow$ Age + 1
---------------------	---------------------------------------	--------------------	--------------------------

Fig. 2. Status update sequence.

time  $t$  is defined as the AoI at the time between status update and new packet arrival, which is in line with the post-action

age definition in [7]. The age of the newly arrived packet is zero, followed by immediate increment at the end of the sequence, and hence the age of a new packet is one at the time of decision.

### III. DECOUPLED MODEL AND INDEX POLICY

In this section, we first formulate the general status update scheduling problem into a MDP problem which is solved by relative value iterations to give a performance benchmark in the simulation section. Afterwards, we introduce the decoupled model where each terminal is examined separately and develop the Whittle's index policy.

#### A. MDP-Based Problem Formulation

The system state is denoted by

$$\mathcal{S}(t) = \{(a_1(t), d_1(t)), \dots, (a_N(t), d_N(t))\}, \quad (6)$$

where  $d_n(t)$  denotes the difference between the AoI of terminal- $n$  and the age of the packet of terminal  $n$ , i.e.,

$$d_n(t) = h_n(t) - a_n(t). \quad (7)$$

Note that this state definition is equivalent with the one with  $(a_n(t), h_n(t))$ , whereas much more convenient in the following derivations and therefore adopted. Based on the fact that the AoI is the last-updated packet's age, it is clear that  $d_n(t) \geq 0, \forall n, t$ ; in addition, we define a new packet's age as one and hence  $a_n(t) \geq 1, \forall n, t$ . The action space is

$$\mathcal{U} = \{1, \dots, N\}, \quad (8)$$

which denotes the scheduled terminal index. Note that with global information, it is without loss of optimality to only consider work-conserving non-collision policies [8, Definition 1]. The state transition probability for terminal- $n$  when not scheduled is described by

$$\begin{aligned} \Pr\{(a_n, d_n) \rightarrow (a_n + 1, d_n)\} &= 1 - \lambda_n; \\ \Pr\{(a_n, d_n) \rightarrow (1, d_n + a_n)\} &= \lambda_n, \end{aligned} \quad (9)$$

and when scheduled,

$$\begin{aligned} \Pr\{(a_n, d_n) \rightarrow (a_n + 1, 0)\} &= 1 - \lambda_n; \\ \Pr\{(a_n, d_n) \rightarrow (1, a_n)\} &= \lambda_n, \end{aligned} \quad (10)$$

with  $a_n \geq 1$  and  $d_n \geq 0$ . The objective is to minimize the long-time average AoI, i.e.,

$$\min_{\pi \in \mathcal{U}} \limsup_{\tau \rightarrow \infty} \frac{1}{\tau N} \sum_{t=1}^{\tau} \sum_{n=1}^N \mathbb{E}[a_{n,\pi}(t) + d_{n,\pi}(t)(1 - u_{n,\pi}(t))]. \quad (11)$$

Such an MDP problem can be solved by the relative value (policy) iteration method with average cost function [13], however, the curse of dimensionality and lack of insights limit the effectiveness of the solution. To address this issue, we note that the above problem can be essentially viewed as an RMAB problem [14] whereby each arm represents one terminal and the reward of pulling an arm is the AoI reduction of the corresponding terminal. It is well-known that the Whittle's

index policy is near-optimal for RMAB problems with a large number of arms, in this case terminals [11]. Therefore, we seek for the index policy in the following subsection.

#### B. Decoupled Model

To design the index policy, a decoupled model is formulated where the AoI of each terminal is compared with an arm with constant cost  $m$ . Whittle's methodology is that by comparing with an arm with constant cost, the value of each arm is represented by the minimum cost that makes the pulling decision of the arm equally beneficial. Therefore, each arm can be investigated separately, and hence the complexity of finding a solution decreases from exponential with  $N$  (MDP value iteration) to linear with  $N$ . Mathematically, the Whittle's index is equivalent to solving a relaxed version of the Lagrange dual problem [10]. However, the main challenge is that the Whittle's index policy is only defined for problems that are indexable, meaning that the value of each state of an arm can be fully characterized by the constant cost of index policy; on the other hand, the existence of such indexability is problem-dependent and usually difficult to establish, especially with multi-dimensional system states whereby simple structures of the solution (e.g., threshold-based) do not exist.

Since only one terminal is considered, we omit the terminal index in this subsection. Concretely, the decoupled problem is formulated by adding a constant cost  $m$  whenever the terminal is scheduled, the objective function of the decoupled model is therefore

$$\hat{J}^* = \min_{u(t) \in \{0,1\}} \frac{1}{\tau} \sum_{t=1}^{\tau} \mathbb{E}[a(t) + d(t)(1 - u(t)) + mu(t)]. \quad (12)$$

We consider the long-time average where  $\tau \rightarrow \infty$ . The action is binary, i.e.,  $u(t) = 0$  or  $u(t) = 1$ . This MDP is an average cost problem with infinite horizon and countably infinite state space, and hence the existence of a deterministic and stationary optimal policy is problem-dependent [13]. However, we note that this specific problem can be shown to have an optimal deterministic and stationary policy by checking that the assumptions outlined in [15, Theorem] are met.

Therefore, the steady states exist and the optimal policy can be obtained by solving the Bellman equations to minimize the long-time average AoI. The Bellman equations are given by (since steady states are considered, the time index is omitted)

$$\begin{aligned} f(a, d) + \hat{J}^* = \\ \min \left\{ \begin{array}{l} d + a + (1 - \lambda)f(a + 1, d) + \lambda f(1, d + a), \\ a + m + (1 - \lambda)f(a + 1, 0) + \lambda f(1, a) \end{array} \right\} \end{aligned} \quad (13)$$

where the top equation in the minimization corresponds to idle and the bottom denotes scheduled, the optimal average cost is denoted by  $\hat{J}^*$ , and  $f(a, d)$  is the differential cost-to-go function and we prescribe  $f(1, 0) = 0$ . In what follows, we will solve the above Bellman equations.

*Theorem 1:* Considering the decoupled model, given an auxiliary cost  $m$ , the optimal action with a state  $(a, d)$  is to schedule the terminal when  $d \geq D_a$  and idle otherwise.

Specifically,  $D_1 = \beta$  where  $\beta$  is the unique positive solution to the following equation:

$$\frac{1}{2}\beta^2 + \left(\frac{1}{\lambda} - \frac{1}{2}\right)\beta - m = 0. \quad (14)$$

The rest of the thresholds are given by

$$D_a = \begin{cases} (1 - \lambda + a\lambda)\beta - \lambda \frac{(a-1)a}{2}, & \text{if } 1 \leq a < \beta; \\ \lambda m, & \text{if } a \geq \beta. \end{cases} \quad (15)$$

The optimum average AoI is

$$\hat{J}^* = \frac{1}{\lambda} + \beta. \quad \square \quad (16)$$

*Proof:* The basic methodology to solve the equations is to first assume the structure of the solution and the optimal policy, and then solve the Bellman equations based on the assumptions. Consistency of the solutions and the assumptions should be checked afterwards. Since there is (with mild conditions) a unique solution to the Bellman equations, such a constructive method can work to find the solution. See Appendix A for details. ■

*Remark 1:* Examining the difference between Theorem 1 and [7, Theorem 5], due to the no-buffer assumption in [7], the system states are nearly one-dimensional (age of a packet is either one or none) and hence there is only one AoI threshold; in contrast, there is one threshold for *each*  $a$  (age of the packet at terminal-side) in Theorem 1, making the derivation of the theorem considerably more challenging. On the other hand, we will show based on simulation results that by using only one-packet buffers, the performance gain is evident compared with the no-buffer index policy. ■

The indexability of the index policy can be readily derived based on Theorem 1.

*Definition 1 (Indexability):* Given costs  $m_1$  and  $m_2$ , and the sets of states that the optimal action is to idle are denoted by  $\Pi_{m_1}$  and  $\Pi_{m_2}$  respectively, the problem is indexable if

$$\forall m_1, m_2 \text{ with } m_1 < m_2 \Rightarrow \Pi_{m_1} \subseteq \Pi_{m_2}, \quad (17)$$

and for  $m = 0$ ,  $\Pi_m = \emptyset$ ; for  $m \rightarrow \infty$ ,  $\Pi_m$  is the entire state space.

*Theorem 2 (Indexability):* Consider the decoupled model and the scheduling policy  $\pi_D$  given in Theorem 1, then  $\pi_D$  is indexable. ■

*Proof:* See Appendix B for details. ■

The index for any state is described as follows.

*Theorem 3 (Whittle's Index):* Consider the decoupled model and denote the index by  $m(a, d)$  with state  $(a, d)$ ,

$$m(a, d) = \begin{cases} \frac{1}{2}x^2 + \left(\frac{1}{\lambda} - \frac{1}{2}\right)x, & \text{if } d > \frac{\lambda}{2}a^2 + \left(1 - \frac{\lambda}{2}\right)a; \\ \frac{d}{\lambda}, & \text{otherwise,} \end{cases} \quad (18)$$

where

$$x \triangleq \frac{d + \frac{a(a-1)\lambda}{2}}{1 - \lambda + a\lambda}. \quad \square \quad (19)$$

*Remark 2:* The derivation of the index follows directly from the optimal policy of the decoupled model in Theorem 1, by the reasoning that the index of a state equals the minimum auxiliary cost that makes the scheduling actions of the terminal under the current state equally beneficial. ■

*Remark 3:* The index is a generalization of the previous results in [6] where the index without randomly generated status packets is derived. In their work, the index is (with transmission success probability  $p = 1$ , user weight  $\alpha = 1$  and frame length  $T = 1$  in [6])

$$C(h) = \frac{1}{2}h(h+1), \quad (20)$$

where  $h$  is the AoI of the terminal. Based on (18), with  $\lambda = 1$  and hence the packet age is  $a = 1$ , we obtain

$$m(1, d) = \frac{1}{2}d^2 + \left(\frac{1}{\lambda} - \frac{1}{2}\right)d = \frac{1}{2}d(d+1). \quad (21)$$

The difference between  $h$  in (20) and  $d = h - 1$  in (21) is due to the fact that [6] adopts the pre-action age and we adopt the post-action age; either case does not affect the results much. The derived index also coincides with [7] when  $a = 1$ . Therefore, it is observed that our results are consistent with previous work and generalize to the scenario with random packet arrivals and arbitrary buffering strategy. ■

#### IV. INDEX-PRIORITIZED RANDOM ACCESS POLICY

The index policy derived above clearly requires global information (all the age and AoI information) for scheduling decisions, and hence it is recognized as a centralized scheduling policy which makes it undesirable in wireless uplinks due to signaling overhead concerns. Nonetheless, inspired by the index policy, we describe an index-prioritized random access policy (IPRA) which achieves the same universally near-optimal AoI performance as the index policy, however, with decentralized protocol structure. A key observation is that each terminal can calculate its own index, denoted by  $I_n$ , based on its transmission history and packet arrivals. Thereby, this individual index is mapped to a transmission probability based on a public mapping function which captures the idea that only valuable packets (packets with high index value) are transmitted; a random access (contention) period is hence introduced to resolve possible collisions. The selection of the public mapping function is tricky and we propose to use a single-threshold function which, notwithstanding its simplicity, achieves near-optimal performance based on simulation results. The detail procedure is described in Algorithm 1.

In IPRA, two parameters, i.e., transmission probability  $p$  and index threshold `indexThreshold`, need to be optimized. In our implementation we find that the optimization is usually easy since the performance as a function of these parameters, i.e.,  $f(p, \text{indexThreshold})$ , is quasi-concave (unimodal). Therefore, well-known methods such as [16] can be applied.



---

**Algorithm 1: IPRA**


---

- 1 **Contention period:** For  $n \in \{1, \dots, N\}$
  - 2 **if**  $I_n \geq \text{indexThreshold}$  **then**
  - 3     Terminal- $n$  transmits with probability  $p$ .
  - 4 **else**
  - 5     Terminal- $n$  is idle.
  - 6 **Transmission/Collision frame:**
  - 7 **If** the transmission is successful, the central controller feeds back an ACK; otherwise a NACK is fed back.
  - 8 **Go to the contention period.**
- 

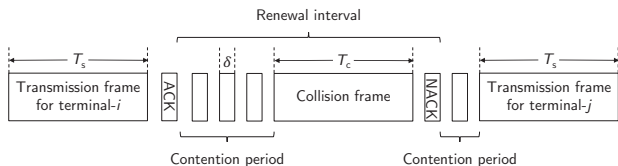


Fig. 3. Transmission timeline of IPRA.

Furthermore, we will show in the following that the overhead of IPRA, which includes the contention period and collision frames, is negligible. Denote the transmission frame length and collision frame length as  $T_s$  and  $T_c$  respectively, and  $\delta$  as the length of one contention slot. We assume  $T_s = T_c$  which is the worst-case assumption meaning that a terminal can only find out the transmission fails after one entire transmission frame.

*Theorem 4:*

$$\lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K T_s \mathbb{1}_{s,k}}{\sum_{k=1}^K \tau_k} \rightarrow 1, \text{ with } \frac{T_s}{\delta} \rightarrow \infty, \quad (22)$$

where  $\tau_k$  denotes the length of the  $k$ -th renewal interval, and  $\mathbb{1}_{s,k}$  is an indicator function denoting whether the transmission is successful in the  $k$ -th renewal interval.  $\square$

*Proof:* The proof is based on the elementary renewal theory and previous work on CSMA in, e.g., [17]. The details are omitted due to lack of space. Moreover, the ratio of  $\frac{T_s}{\delta}$  is usually very high (around 100 is common [17]) and hence the overhead of IPRA is negligible.  $\blacksquare$

## V. SIMULATION RESULTS

In this section, we present numerical results based on computer simulations which run the scheduling policies for  $10^6$  time slots and obtain the time-average AoI. It is no surprise that the Whittle's index policy is near-optimal given many existing works which have validated this conclusion in various scenarios. In view of this, the main purpose of the section is to highlight the performance difference between the Whittle's index policy with optimal packet management (one-packet buffers) and several existing scheduling policies [7], [8], and to demonstrate the performance of IPRA.

First in Fig. 4, a 2-terminal case is considered and the optimum AoI based on solving the MDP via relative value

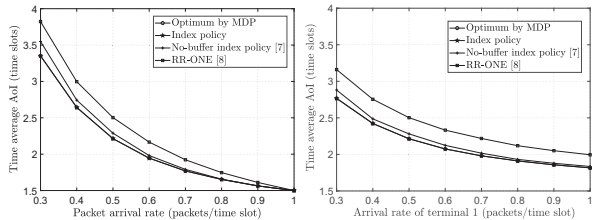


Fig. 4. Performance comparisons with optimum by solving the MDP and two terminals having identical packet arrival rate (left) and heterogeneous arrival rates (right) with terminal 1's arrival rate shown as x-axis and terminal 2's fixed as 0.5.

iteration is obtained. The difference between the Whittle's index policy and the optimum is hardly visible, whereas the no-buffer index policy suffers evident performance loss when the packet arrival rate is low. This can be explained that the no-buffer strategy drops precious (when with low arrival rates) status packets when both terminals have arrivals in the same time slot; in this case packets should be stored for update in the future. Note that the no-buffer index policy is identical with the index policy in this paper when  $\lambda$  approaches one, by showing that the index in Theorem 2 in this regime, i.e., (21), coincides with the index in [7].<sup>1</sup>

The performance gap between one-buffer and no-buffer index policies is more pronounced in Fig. 5, where we also simulate the decentralized policy RR-ONE in [8] and IPRA. It has been shown in [8] that RR-ONE achieves the optimal scaling factor with a large  $N$  but is suboptimal with finite  $N$ ; this is observed in the figure. It is shown that the gap between no-buffer and one-buffer policies is most evident when the mean arrival interval is comparable with the number of terminals, i.e., the joint-asymptotic regime where  $N \rightarrow \infty$ ,  $\frac{1}{\lambda} \rightarrow \infty$ , and  $N\lambda \rightarrow C$ , where  $C$  is a fixed constant. This is because in this regime, the delays due to scheduling among terminals and random packet arrivals are equally significant and hence neither can be ignored, making the optimal policy most elusive; when  $N \rightarrow \infty$  the delay due to random packet arrivals can be neglected and hence dropping packets by no-buffer policy is also near-optimal.

More importantly, we find in Fig. 5 that IPRA achieves universally near-optimal performance by noting that its performance is very close to the one-buffer index policy. The length of one contention slot is set to be  $1/100$  [17] of a transmission frame which is of the same length of a time slot in RR-ONE and index policies. The transmission probability and index threshold in IPRA are optimized based on a bi-section search algorithm.

To demonstrate the benefit of buffering packets over no-buffer policies, we compare the normalized indices with various  $a$  in Fig. 6; note that only the relative value of the index is relevant hence justifying the normalization. The benefit of buffering is shown by the fact that the index value of a state with the packet's age larger than one is still significant;

<sup>1</sup>The difference between  $h$  here and  $h + 1$  in [7] is due to the difference in the definitions of initial packet age.

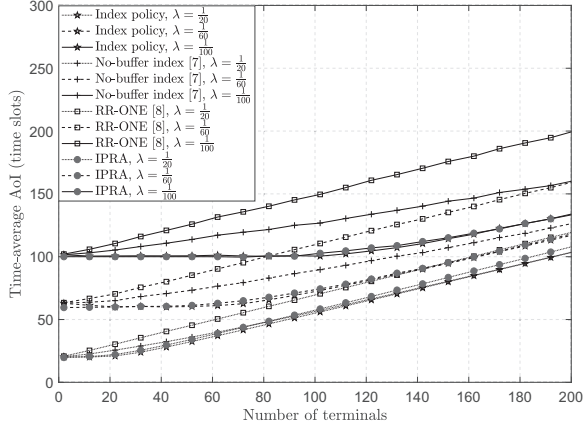


Fig. 5. Performance comparisons with a large number of terminals with identical packet arrival rate of  $\lambda$  specified in the legend.

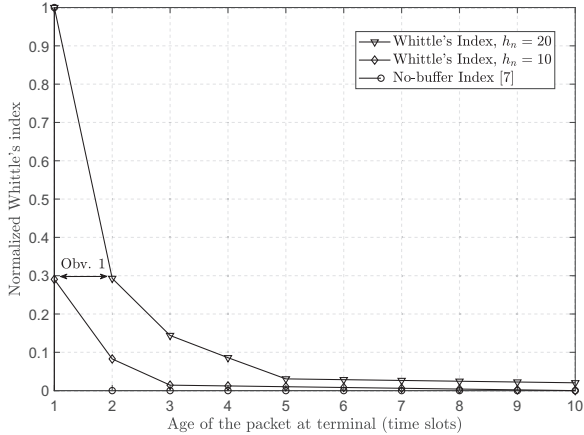


Fig. 6. Illustration of the difference between the Whittle's indices with optimal buffering strategy and no-buffer.

hence dropping the packet of such kind losses performance. In particular, based on observation 1 in the figure, a state with  $(a, h) = (1, 10)$  has approximately the same index with a state  $(2, 20)$ , meaning that scheduling a terminal with state  $(2, 20)$  (buffering the packet for one time slot) is equally beneficial with one with  $(1, 10)$ , whereas the no-buffer index deems the state  $(2, 20)$  as zero-value which is clearly sub-optimal.

## VI. CONCLUSIONS

In summary, the answer to the question whether decentralized status update can achieve universally near-optimal AoI in wireless multiaccess channels is positive based on the following presented results: (a) The centralized Whittle's index policy which has near-optimal performance is derived in closed-form and the indexability is established. The performance thereof serves as a performance benchmark for optimality. (b) IPRA adopts the derived index to prioritize packet transmissions among terminals and a random access procedure is applied with proven negligible overhead, such

that the status update is decentralized. Moreover, it is shown that the performance of IPRA is nearly identical with the index policy and outperforms state-of-the-arts in the literature.

For future IoT applications where the main objective is to optimize the status update timeliness in wireless uplinks, IPRA is a promising decentralized multiple-access solution achieving better information freshness with smaller signaling overhead compared with current approaches, e.g., CSMA and grant-based schemes in cellular systems. Future work includes considering arbitrary packet arrival patterns such as bursty arrivals, as well as non-Markovian information sources.

## ACKNOWLEDGEMENT

This work is sponsored in part by the Nature Science Foundation of China (No. 61701275, No. 91638204, No. 61571265, No. 61621091), the China Postdoctoral Science Foundation, and Hitachi Ltd.

## APPENDIX A

### PROOF FOR THEOREM 1

As a first step, given an arbitrary auxiliary cost  $m \geq 0$ , two conditions regarding the optimal policy and the solution are assumed and later verified.

*Condition 1:* The optimal policy is a threshold policy, denoted by  $\pi_D$ , that is given any state  $(a, d)$ , the optimal action is to idle when  $0 \leq d < D_a$  and to schedule when  $d \geq D_a$ . Furthermore, the thresholds satisfy

$$D_1 \leq D_2 \leq \dots \leq D_a \leq \dots \quad \square \quad (23)$$

*Condition 2:* For any  $a$ ,

$$f(a, 0) \leq f(a, 1) \leq \dots \leq f(a, d) \leq \dots \quad \square \quad (24)$$

Both conditions are in fact intuitive as explained in the following. The threshold-based policy structure in Condition 1 stems from the tradeoff between auxiliary cost and AoI reduction ( $d$ ) in each time slot; the monotonicity of the thresholds in (23) reflects the fact that with a relatively older packet (larger  $a$ ) at terminal-side, it is better wait for a new packet arrival than to schedule this time. The Condition 2 simply states the fact that a state with a larger AoI and the same packet age has a larger cost-to-go value. Although reasonable, after we solve the Bellman equations, we will check the consistency of these assumptions. For ease of exposition, we assume the thresholds can be real values in this section.

With  $\pi_D$ , an important property regarding the differential cost-to-go function is stated below.

*Proposition 1:* For any  $a_1, a_2 \geq 1$ ,  $0 \leq d_1 < D_{a_1}$ ,  $0 \leq d_2 < D_{a_2}$ , and  $a_1 + d_1 = a_2 + d_2$ , we have

$$f(a_1, d_1) = f(a_2, d_2). \quad \square \quad (25)$$

*Proof:* To prove this proposition, we first prove the following lemma.

*Lemma 1:* There exists  $a_M$  such that  $\forall a \geq a_M$ , the threshold  $D_a = D_{a_M}$  is a constant.  $\square$

*Proof:* Based on (23), the thresholds are monotonically increasing with respect to  $a$ . Note that when  $d \geq m$ , observing

the cost-to-go equations in (13), we can show that the optimal action in this case is to schedule the terminal, that is the upper part in the minimization (denoted by  $\mu_0(a, d)$ ) is larger than the lower part (denoted by  $\mu_1(a, d)$ ). Concretely, with  $d \geq m$ ,

$$\begin{aligned} & \mu_0(a, d) - \mu_1(a, d) \\ &= d - m + (1 - \lambda)(f(a + 1, d) - f(a + 1, 0)) \\ & \quad + \lambda(f(1, d + a) - f(1, a)) \geq 0. \end{aligned} \quad (26)$$

The equality is based on Condition 2. This means that  $\forall a, D_a \leq m$ . Therefore, the threshold array  $\{D_a : a = 1, 2, \dots\}$  is monotonically increasing with a finite upper bound; hence the limitation exists and this concludes the proof. ■

Given  $\pi_D$  and based on Lemma 1, the action when  $a \geq a_M$  and  $d \geq D_{a_M}$  is to schedule the terminal. Therefore, based on (13) and Condition 1 we have  $D_{a_M} \geq D_1$ , and hence

$$f(1, d) + \hat{J}^* = 1 + m + (1 - \lambda)f(2, 0) + \lambda f(1, 1), \quad d \geq D_{a_M}. \quad (27)$$

Note that based on (27),  $f(1, d) = f(1, D_{a_M})$ ,  $\forall d \geq D_{a_M}$ . Additionally, the action when  $d = D_{a_M} - 1$  and  $a \geq a_M$  is to idle, and hence,

$$\begin{aligned} & f(a, D_{a_M} - 1) + \hat{J}^* \\ &= D_{a_M} + a - 1 + (1 - \lambda)f(a + 1, D_{a_M} - 1) \\ & \quad + \lambda f(1, D_{a_M} + a - 1) \\ &= D_{a_M} + a - 1 + (1 - \lambda)f(a + 1, D_{a_M} - 1) \\ & \quad + \lambda f(1, D_{a_M}). \end{aligned} \quad (28)$$

Denote

$$g \triangleq D_{a_M} - \hat{J}^* - 1 + \lambda f(1, D_{a_M}), \quad (29)$$

we obtain

$$f(a, D_{a_M} - 1) = a + (1 - \lambda)f(a + 1, D_{a_M} - 1) + g. \quad (30)$$

Solving (30) recursively with respect to  $a$  yields

$$f(a, D_{a_M} - 1) = \gamma_0(1 - \lambda)^{a_M - a} + \frac{a + g}{\lambda} + \frac{1 - \lambda}{\lambda^2} \quad (31)$$

where  $\gamma_0 = f(a_M, D_{a_M} - 1) - \frac{a_M + g}{\lambda} - \frac{1 - \lambda}{\lambda^2}$  and  $a \geq a_M$ . When  $\gamma_0 \neq 0$ , the expression has an exponential term which is unreasonable because this means a state with a larger packet age (and the same  $d$ ) has an exponentially larger cost; moreover, we will show the solution with  $\gamma_0 = 0$  is a valid solution by checking the consistency. A condition is used in [18] under similar circumstances by claiming

$$\lim_{a \rightarrow \infty} (1 - \lambda)^a f(a, D_{a_M} - 1) = 0, \quad (32)$$

which essentially carries the same meaning as ours. Likewise, we obtain

$$\begin{aligned} f(a, D_{a_M} - s) &= \frac{a + g - s}{\lambda} + \frac{1}{\lambda^2}, \\ a &\geq s + a_M - 1, \quad 1 \leq s \leq D_{a_M}, \end{aligned} \quad (33)$$

Note that, e.g.,

$$\begin{aligned} f(a_M, D_{a_M} - 1) &= f(a_M + 1, D_{a_M} - 2) = \dots \\ &= f(D_{a_M} + a_M - 1, 0). \end{aligned} \quad (34)$$

It is therefore clear that the proposition holds for any  $a_1, a_2, d_1, d_2$  satisfying  $a_1 + d_1 = a_2 + d_2 \geq D_{a_M} + a_M - 1$ . For  $a_1, a_2, d_1, d_2$  satisfying  $a_1 + d_1 = a_2 + d_2 < D_{a_M} + a_M - 1$ , an induction based proof is adopted.

Suppose the proposition holds for any  $a_1, a_2, d_1, d_2$  satisfying  $a_1 + d_1 = a_2 + d_2 = D_{a_M} + a_M - 1 - s$ ,  $0 \leq s \leq D_{a_M} + a_M - 2$ ,  $d_1 < D_{a_1}$  and  $d_2 < D_{a_2}$ , then for any  $a'_1, a'_2, d'_1, d'_2$  satisfying  $a'_1 + d'_1 = a'_2 + d'_2 = D_{a_M} + a_M - 2 - s$  and  $d'_1 < D_{a'_1}$ ,  $d'_2 < D_{a'_2}$ , the action is to idle based on  $\pi_D$ . It follows from (13) that

$$\begin{aligned} & f(a'_1, d'_1) \\ &= -\hat{J}^* + d'_1 + a'_1 + (1 - \lambda)f(a'_1 + 1, d'_1) + \lambda f(1, d'_1 + a'_1) \\ & \stackrel{(a)}{=} -\hat{J}^* + d'_2 + a'_2 + (1 - \lambda)f(a'_2 + 1, d'_2) + \lambda f(1, d'_2 + a'_2) \\ &= f(a'_2, d'_2). \end{aligned} \quad (35)$$

The equality (a) is based on the induction hypothesis. Also note that  $d'_1 < D_{a'_1} \leq D_{a'_1 + 1}$  based on Condition 1, and hence the conditions are all satisfied. For the induction basis, the proposition holds for  $s = 0$  based on (33). Therefore, the proposition is concluded. ■

For  $1 \leq a < D_1$ , based on (13),

$$\begin{aligned} f(a, 0) &= -\hat{J}^* + a + (1 - \lambda)f(a + 1, 0) + \lambda f(1, a) \\ &= -\hat{J}^* + a + f(a + 1, 0), \end{aligned} \quad (36)$$

where the last equality is based on Proposition 1. Given that  $f(1, 0) = 0$ , it follows that

$$f(a, 0) = (a - 1)\hat{J}^* - \frac{a(a - 1)}{2}, \quad 1 \leq a \leq D_1. \quad (37)$$

To proceed, we obtain another important property of the solution in the following proposition.

*Proposition 2:* For any state  $(a, d)$  with  $a \geq 1$  and  $d \geq D_a$ ,

$$f(a, d) - f(a, 0) = m. \quad (38)$$

*Proof:* For some state  $(a, d)$  with  $a \geq 1$  and  $d \geq D_a$ , based on (13) and  $\pi_D$ ,

$$\begin{aligned} f(a, d) &= -\hat{J}^* + a + m + (1 - \lambda)f(a + 1, 0) + \lambda f(1, a) \\ f(a, 0) &= -\hat{J}^* + a + (1 - \lambda)f(a + 1, 0) + \lambda f(1, a). \end{aligned} \quad (39)$$

Therefore the proposition is concluded by observing the difference of the two above equations. ■

*Remark 4:* This proposition is reasonable since for states  $(a, d)$  with  $d \geq D_a$  and  $\pi_{(D)}$ , the relative cost compared with state  $(a, 0)$  is a one-time scheduling auxiliary cost  $m$ . A direct corollary is that

$$f(1, d) = f(1, 0) + m = m, \quad d \geq D_1. \quad \square \quad (40)$$

Based on Proposition 1 and the same arguments in (31), we obtain

$$f(a, 0) = \frac{a - \hat{J}^* - 1}{\lambda} + m + \frac{1}{\lambda^2}, \quad a \geq D_1. \quad (41)$$

Based on Proposition 2, when  $d \geq D_a$  it follows that

$$\begin{aligned} & f(a, d) \\ &= m + f(a, 0) \\ &= \begin{cases} m + a\hat{J}^* - \frac{(a-1)a}{2}, & \text{if } 1 \leq a < D_1; \\ \frac{a}{\lambda} + 2m - \frac{1-\lambda}{\lambda}\hat{J}^* + \frac{1-\lambda}{\lambda^2}, & \text{if } a \geq D_1. \end{cases} \end{aligned} \quad (42)$$

Combining (37) and (41) when  $d = D_1$  gives us the relationship among  $m$ ,  $\hat{J}^*$  and  $D_1$ , i.e.,

$$m = \left(D_1 - 1 + \frac{1}{\lambda}\right)\hat{J}^* - \frac{D_1^2}{2} + \frac{D_1}{2} - \frac{D_1}{\lambda} + \frac{\lambda-1}{\lambda^2}. \quad (43)$$

Now we have obtained all differential cost-to-go function expressions based on  $\pi_D$ . To show that this is the solution to the Bellman equations, we need to demonstrate the policy  $\pi_D$ , whereby we assume the optimal policy has certain properties in Condition 1 and 2, is indeed consistent. In this regard, it will be shown that by carefully arranging the thresholds  $D_a$  with  $a = 1, 2, \dots$ , the threshold policy  $\pi_D$  is the solution to the Bellman equations in each state. Let us first check the boundary points, i.e., state  $(a, D_a)$ .

For state  $(a, D_a)$ , observe the cost-to-go equations in (13), we need to show that the optimal action in this case is to schedule the terminal, that is the upper part in the minimization (denoted by  $\mu_0(a, D_a)$ ) is larger than the lower part (denoted by  $\mu_1(a, D_a)$ ). First, consider the case where  $a \geq D_1$ , the difference is

$$\begin{aligned} & \mu_0(a, D_a) - \mu_1(a, D_a) \\ &= D_a - m + (1-\lambda)(f(a+1, D_a) - f(a+1, 0)) \\ & \quad + \lambda(f(1, D_a + a) - f(1, a)) \\ &\stackrel{(a)}{=} D_a - m + (1-\lambda)\frac{D_a}{\lambda} \geq 0, \end{aligned} \quad (44)$$

where the equality (a) is attributed to (40) and (41). Therefore we obtain a condition

$$D_a \geq \lambda m. \quad (45)$$

The case with  $1 \leq a < D_1$  yields

$$\begin{aligned} & \mu_0(a, D_a) - \mu_1(a, D_a) \\ &= D_a - m + (1-\lambda)(f(a+1, D_a) - f(a+1, 0)) \\ & \quad + \lambda(f(1, D_a + a) - f(1, a)) \\ &\stackrel{(a)}{=} D_a - m + (1-\lambda)f(a+1, D_a) + \lambda m - f(a+1, 0) \\ &\stackrel{(b)}{=} D_a - m + (1-\lambda)\left(\frac{D_a + a - \hat{J}^*}{\lambda} + m + \frac{1}{\lambda^2}\right) \\ & \quad + \lambda m - a\hat{J}^* + \frac{a(a+1)}{2} \geq 0, \end{aligned} \quad (46)$$

where the equality (a) follows from Proposition 1 and 2 and equality (b) is from (41). We obtain the second condition:

$$D_a \geq (1-\lambda+a\lambda)\hat{J}^* - a + 1 - \lambda\frac{a(a-1)}{2} - \frac{1}{\lambda}. \quad (47)$$

Furthermore, for some state  $(a, d)$  with  $d > D_a$ , based on Condition 2 and (13) the upper part which is associated with idling the terminal keeps increasing while the lower part stays the same, it is clear that

$$\mu_0(a, d) - \mu_1(a, d) \geq 0, \quad d > D_a, \quad (48)$$

once (45) and (47) are both satisfied.

Then the states are examined in which, based on  $\pi_D$ , the terminal should be idle. Specifically, for some state  $(a, D_a - 1)$  with  $a \geq D_1$ , we obtain

$$\begin{aligned} & \mu_0(a, D_a - 1) - \mu_1(a, D_a - 1) \\ &= D_a - 1 - m + (1-\lambda)(f(a+1, D_a - 1) - f(a+1, 0)) \\ & \quad + \lambda(f(1, D_a - 1 + a) - f(1, a)) \\ &\stackrel{(a)}{=} D_a - 1 - m + \frac{1-\lambda}{\lambda}(D_a - 1) \leq 0, \end{aligned} \quad (49)$$

where the equality (a) follows from (41), and thus another condition is

$$D_a \leq \lambda m + 1. \quad (50)$$

With  $1 \leq a < D_1$ ,

$$\begin{aligned} & \mu_0(a, D_a - 1) - \mu_1(a, D_a - 1) \\ &= D_a - 1 - m + (1-\lambda)(f(a+1, D_a - 1) - f(a+1, 0)) \\ & \quad + \lambda(f(1, D_a - 1 + a) - f(1, a)) \\ &\stackrel{(a)}{=} D_a - 1 - m + (1-\lambda)f(a+1, D_a - 1) \\ & \quad + \lambda m - f(a+1, 0) \\ &\stackrel{(b)}{=} D_a - 1 - m + (1-\lambda)\left(\frac{D_a + a - \hat{J}^* - 1}{\lambda} + m + \frac{1}{\lambda^2}\right) \\ & \quad + \lambda m - a\hat{J}^* + \frac{a(a+1)}{2} \leq 0, \end{aligned} \quad (51)$$

where the equality (a) follows from Proposition 1 and 2 and equality (b) is from (41). Hence, the following condition should be satisfied.

$$D_a \leq (1-\lambda+a\lambda)\hat{J}^* - a + 2 - \lambda\frac{a(a-1)}{2} - \frac{1}{\lambda}. \quad (52)$$

Following the same arguments before, for some state  $(a, d)$  with  $d < D_a - 1$ , based on Condition 2 and (13) the upper part which is associated with idling the terminal keeps decreasing when  $d$  decreases while the lower part stays the same, it is thus clear that

$$\mu_0(a, d) - \mu_1(a, d) \leq 0, \quad d < D_a - 1, \quad (53)$$

once (50) and (52) are both satisfied. To summarize, the valid solution should satisfy the following conditions as well as (43).

$$\begin{aligned} & (1-\lambda+a\lambda)\hat{J}^* - a + 1 - \lambda\frac{a(a-1)}{2} - \frac{1}{\lambda} \leq D_a \leq \dots \\ & (1-\lambda+a\lambda)\hat{J}^* - a + 2 - \lambda\frac{a(a-1)}{2} - \frac{1}{\lambda}, \text{ if } 1 \leq a < D_1; \\ & \lambda m \leq D_a \leq \lambda m + 1, \text{ if } a \geq D_1. \end{aligned} \quad (54)$$



Solving for the solution to the above, we obtain  $D_1$  is the unique positive solution to the following equation:

$$\frac{1}{2}D_1^2 + \left(\frac{1}{\lambda} - \frac{1}{2}\right)D_1 - m = 0. \quad (55)$$

The uniqueness of the positive solution can be shown based on the above formula. The rest of the thresholds are given by

$$D_a = \begin{cases} (1 - \lambda + a\lambda)D_1 - \lambda \frac{(a-1)a}{2}, & \text{if } 1 \leq a < D_1; \\ \lambda m, & \text{if } a \geq D_1. \end{cases} \quad (56)$$

The average AoI is

$$\hat{J}^* = \frac{1}{\lambda} + D_1. \quad (57)$$

Thus far, we have obtained the solution based on Condition 1 and 2. What is still left to do is to show that the solution is in consistency with both assumptions. Therefore, we first check the first assumption. The threshold policy nature has been validated in the preceding proofs, i.e., the optimal action given these differential cost-to-go functions is indeed threshold-based. Moreover, the monotonicity of the thresholds is shown as follows. Observe (56), for  $1 \leq a < D_1$

$$D_{a+1} - D_a = \lambda(D_1 - a) > 0. \quad (58)$$

It follows that

$$D_1 < D_2 < \dots < D_{D_1-1}. \quad (59)$$

Moreover,

$$\begin{aligned} D_{D_1-1} - D_{a, a \geq D_1} &\stackrel{(a)}{\leq} D_1^2 - \lambda \frac{(D_1-1)D_1}{2} - \lambda m \\ &\stackrel{(b)}{=} \left(2 - \frac{1}{\lambda} - \lambda\right)D_1 + \lambda - 2 + \frac{1}{\lambda} \\ &= 0, \end{aligned} \quad (60)$$

wherein the equality (a) follows from the fact that  $D_{D_1-1} \leq D_{D_1}$  based on (58), and the equality (b) is based on (55) and massaging the terms. With this, we have shown that the obtained solution is consistent with Condition 1.

For Condition 2 and specifically (24), when  $a + d < D_1$ ,

$$f(a, d+1) - f(a, d) \stackrel{(a)}{=} \hat{J}^* - a \stackrel{(b)}{>} D_1 + \frac{1}{\lambda} - a > 0, \quad (61)$$

where the inequality (a) follows from (37) and (b) is from (57). When  $D_1 - a \leq d < D_a$ , the monotonicity is ensured by

$$f(a, d+1) - f(a, d) \stackrel{(a)}{=} \frac{1}{\lambda} > 0, \quad (62)$$

where the equality (a) is based on Proposition 1 and (41). Note that the monotonicity around the connection point  $(a, D_1 - a)$  is intact by observing (43). Furthermore, when  $d \geq D_a$ ,  $f(a, d)$  is a constant with respect to  $d$  which is shown by Proposition 2 and thus the monotonicity also holds. The connection point  $(a, D_a)$  is validated by the same arguments as in (60). We conclude the proof with this.

## APPENDIX B

### PROOF FOR THEOREM 2

Observe the thresholds in Theorem (1). With  $m = 0$ , the maximum threshold equals zero. Moreover, based on the monotonicity of the thresholds which is shown in the proof of Theorem 1, all thresholds are zero and hence the idle state space is an empty space. On the other hand, when  $m$  goes to infinity, all the thresholds go to infinity with it and hence the idle state space approaches the entire space. Moreover, for any  $m_1 < m_2$ , the thresholds associated with  $m_1$  are no larger than those with  $m_2$ , and hence the monotonicity condition follows straightforwardly. This concludes the proof.

### REFERENCES

- [1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM*, Mar 2012, pp. 2731–2735.
- [2] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *IEEE Int'l Symp. Info. Theory*, Jun 2015, pp. 1681–1685.
- [3] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inform. Theory*, vol. 62, no. 4, pp. 1897–1910, April 2016.
- [4] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *IEEE Int'l Symp. Info. Theory*, Jun 2017, pp. 131–135.
- [5] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," in *IEEE INFOCOM*, April 2016, pp. 1–9.
- [6] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Annu. Allerton Conf. Commun., Control, Comput.*, Sep 2016, pp. 844–851.
- [7] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," *arXiv preprint arXiv:1801.03422*, 2018.
- [8] 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.
- [9] Online; accessed 20-Jan-2018; <http://www.auburn.edu/~yzs0078/>.
- [10] W. Whitt, "Minimizing delays in the GI/G/1 queue," *Operations Research*, vol. 32, no. 1, pp. 41–51, 1984.
- [11] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, vol. 27, no. 3, p. 637–648, 1990.
- [12] E. Najm and R. Nasser, "The age of information: The gamma awakening," in *IEEE Int'l Symp. Info. Theory*, 2016, pp. 2574–2578.
- [13] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1.
- [14] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: General framework, results and applications," in *International Conference on Machine Learning*, 2013, pp. 1–151–1–159.
- [15] L. I. Sennott, "Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs," *Operations Research*, vol. 37, no. 4, pp. 626–633, 1989.
- [16] R. Combes and A. Proutiere, "Unimodal bandits: Regret lower bounds and optimal algorithms," in *Proc. International Conference on Machine Learning*, vol. 32, no. 1, 2014, pp. 521–529.
- [17] Y. Gai, S. Ganesan, and B. Krishnamachari, "The saturation throughput region of p-persistent CSMA," in *Information Theory and Applications Workshop*, Feb 2011, pp. 1–4.
- [18] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *arXiv preprint arXiv:1801.01803*, 2018.