Optimizing Age of Information in Wireless Uplink Networks With Partial Observations

Jingwei Liu, Rui Zhang, Aoyu Gong, and He Chen, *Member, IEEE*

***Abstract*— This paper considers a wireless uplink network consisting of multiple end devices and an access point (AP). Each device monitors a physical process with randomly generated status updates and sends these update packets to the AP in the uplink. The AP aims to schedule the transmissions of these devices to optimize the network-wide information freshness, quantified by the age of information (AoI) metric. Due to the stochastic arrival of the status updates at end devices, the AP only has *partial observations* of system times of the latest status update packets at end devices when making scheduling decisions. Such a decision-making problem can be naturally formulated as a partially observable Markov decision process (POMDP). We reformulate the POMDP into an equivalent belief Markov decision process (belief-MDP), by defining fully observable belief states of the POMDP as the states of the belief-MDP. The belief-MDP in its original form is difficult to solve as the dimension of its states can go to infinity and its belief space is uncountable. Fortunately, by carefully leveraging the properties of the status update arrival processes (i.e., Bernoulli processes), we manage to simplify the belief-MDP substantially, where every feasible state is characterized by a two-dimensional vector. Based on the simplified belief-MDP, we devise a low-complexity scheduling policy, termed Partially Observing Max-Weight (POMW) policy, for the formulated AoI-oriented scheduling problem. We derive upper bounds for the time-average AoI performance of the proposed POMW policy. We analyze the performance guarantee for the POMW policy by comparing its performance with a universal lower bound available in the literature. Numerical results validate our analyses and demonstrate that the performance gap between the POMW policy and its fully observable counterpart is proportional to the inverse of the lowest arrival rate of all end devices.**

***Index Terms*— Age of information, multiuser scheduling, partially observable Markov decision process, belief Markov decision process.**

Manuscript received 27 September 2022; revised 6 February 2023; accepted 30 March 2023. Date of publication 6 April 2023; date of current version 17 July 2023. The work of Jingwei Liu and He Chen are supported in part by the Innovation and Technolgy Fund (ITF) under Project ITS/204/20 and the CUHK direct grant for research under Project 4055166. The work of Rui Zhang is supported in part by the Research Talent Hub PiH/380/21 under Project ITS/204/20. An earlier version of this paper was presented in part at the IEEE GLOBECOM 2020 [DOI: 10.1109/GLOBECOM42002.2020.9348022]. The associate editor coordinating the review of this article and approving it for publication was C. Jiang. *(Corresponding author: He Chen.)*

Jingwei Liu, Rui Zhang, and He Chen are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China (e-mail: lj020@ie.cuhk.edu.hk; ruizhang@ie.cuhk.edu.hk; he.chen@ie.cuhk.edu.hk).

Aoyu Gong is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: aoyu.gong@epfl.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2023.3265091.

Digital Object Identifier 10.1109/TCOMM.2023.3265091

I. Introduction

THE rapid development of wireless communication technologies in the past decades has stimulated their ubiquitous applications in time-critical systems, such as vehicular networks and industrial control networks [1], [2], [3]. In these applications, information (e.g., velocity and position of a vehicle) needs to be delivered to targeted receivers as timely as possible. The stale information could cause severe consequences, e.g., damages to facilities or even losses of human lives. Hence, the information timeliness or freshness in these networks is of great importance. To quantify the information freshness, the age of information (AoI) metric has been proposed and extensively investigated in the literature (e.g., see [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] and references therein). More specifically, AoI is defined as the time elapsed since the generation of the last successfully received message at destination [4]. Many efforts have been made on tackling transmission scheduling problems to minimize the time-average AoI of various network settings. Early work focused on the AoI-based transmission scheduling problem in single-user networks, see e.g., [20], [21], [22], [23], [24], [25], where the AoI performance of the single user was optimized by determining when to transmit a status update packet. Recent work has shifted to design the AoI-based scheduling policies for multiuser networks, see e.g., [26], [27], [28], [29], [30], [31], [32]. In these work, the network-wide time-average AoI was optimized by determining how to schedule the transmission sequence of multiple users.

In downlink multiuser networks, an access point (AP) monitors multiple information sources and schedules transmissions of the generated status update packets from itself to the corresponding end devices, respectively. In this context, the AP can completely know the evolution of AoI when acknowledgements are provided by end devices. The AoI-based scheduling problems in downlink multiuser networks were thoroughly studied in [26], [28], and [33]. The authors in [26] considered the "generate-at-will" model for the generation of status updates. In this model, the AP generates a status update for an information source whenever the transmission to its targeted end device is scheduled. As such, the AP only needs to consider the instantaneous AoI values of all end devices when making scheduling decisions. Authors in [26] first proved that in symmetric networks, a greedy policy, which schedules the end device with the highest value of instantaneous AoI, is optimal for minimizing the long-term

average AoI. For more general networks, three low-complexity scheduling policies were proposed and compared, including a Max-Weight policy derived from the Lyapunov optimization framework [34], a randomized policy, and a Whittleâ's Index policy. Reference [28] extended the Max-Weight policy to the downlink networks with the "stochastic arrival" model, and an upper bound for the network-wide time-average AoI was derived. On the other hand, [27] developed a Whittle's Index policy for the same scenario as in [28]. In the "stochastic arrival" model, the generation of status update packets for each information source follows a stochastic process. In this case, the system times of update packets at the AP and the instantaneous AoI values of all end devices need to be jointly considered when designing the scheduling policies for the AP.

In uplink multiuser networks, on the other hand, each end device monitors the statuses of a separate information source and sends status update packets to a common AP. The AP aims to maintain a low network-wide AoI performance by carefully scheduling the transmissions of status update packets in the uplink. As the information destination, the AP has a full track of the AoI values of all streams of status updates. For the "generate-at-will" model, each node will generate a new status update packet once granted to transmit. As such, the system times of status update packets are fixed and can be perfectly known to the AP. In this case, the AoI-oriented scheduling problems in uplink networks are mathematically equivalent to those in downlink networks when the scheduling constraints of the two types of networks are the same. By contrast, when it comes to the "stochastic arrival" model, the scheduling problems in uplink multiuser networks are largely different from those in downlink networks. This is because in uplink networks, the AP may need to make scheduling decisions under partial observations of the system times of randomly generated status update packets at end device side. The complete observations of the system times of all status update packets requires end devices to report the arrivals of new status updates to the AP before each scheduling decision-making. Such a reporting procedure could lead to non-negligible network overhead, especially when status update packets are short. Therefore, it is of practical significance to address a question that arises here: Can we still effectively optimize the AoI performance even if the status update arrivals at the end nodes are not reported to the AP? In that context, the AP only has an observation of the system time of status update of a certain end device only when the device is scheduled to transmit and the transmission is successful. To the best knowledge, such an AoI-based scheduling problem for uplink multiuser networks with partial observations has not been thoroughly studied in open literature. We note that [27] developed a Whittle's Index policy for optimizing AoI in an uplink multiuser network with the "stochastic arrival" model. However, the system times of status update packets at all nodes are assumed to be fully observed, making the scheduling problem mathematically equivalent to that in [28].

As an attempt to fill the gap, in this paper we aim to optimize the *expected weighted sum AoI* for an uplink multiuser network with stochastic arrivals of status updates at end devices. The arrivals of status update packets at end devices are assumed to follow independent Bernoulli processes, which is commonly used in the literature (see e.g., [26], [27], [28], [34]). We consider that the end devices will not report the random arrivals of the status updates to the AP for minimizing the network overhead. As such, the designed scheduling policy needs to make decisions with partial observations. The main contributions of this paper are summarized as follows.

- We formulate our AoI-oriented scheduling problem as a POMDP problem considering the incomplete knowledge of status update arrivals of end devices at the AP. The instantaneous system times of status update packets at the end devices and the instantaneous AoI at the AP are jointly defined as the states of the POMDP. We reformulate the POMDP to an equivalent belief Markov decision process (belief-MDP), where the states of the belief–MDP, termed belief states, are defined as the posterior distributions of the states of the POMDP. We remark that computing the optimal policy for the belief-MDP (or the POMDP) is a PSPACE-complete problem [35], which is not practically computable. Nevertheless, such a belief-MDP reformulation benefits the policy design and the theoretical analysis since the belief states characterize sufficient statistics of the system.
- To solve the formulated belief-MDP, we propose an effective simplification to characterize all feasible infinite-dimensional belief states as two-dimensional vectors. This is achieved by analyzing how Bernoulli arrival processes of status updates at end devices affect the evolution of the belief states. By doing so, we reduce the continuous spaces of the belief states to discrete ones. That is, we extract the feasible belief spaces from the corresponding distribution spaces. The simplification of belief updates in belief-MDP largely facilitate the design of scheduling policies as well as the theoretical analysis of the scheduling policies' performance.
- We devise a low-complexity Partially Observable Max-Weight (POMW) policy, inspired by the Lyapunov optimization framework [34]. The POMW policy aims to minimize a Lyapunov Drift function, defined as the expectation of the sum of weighted instantaneous AoI, in each time slot under condition of the current belief states. Based on the simplified belief-MDP model and a Randomized Scheduling policy proposed in [28], we derive upper bounds for the expected weighted sum AoI performance of the POMW policy. Further, we evaluate the performance guarantee for the POMW policy, which is defined as the ratio between the AoI performance of the POMW policy and that of a universal lower bound. Simulation results validate our theoretical analysis. Simulation results also show that the performance gap between the POMW policy and its fully observable counterpart is inversely proportional to the lowest arrival rate of all end devices. Moreover, the proposed POMW policy is superior to the baseline policies, which do not use the statistical information of the system times of the status update packets at end devices.

We notice a handful of efforts on designing AoI-oriented scheduling policies that also considered networks with partial

observations [21], [36], [37], [38]. Leng and Yener investigated the AoI minimization in a time-slotted cognitive radio energy harvesting network [21]. In [21], a secondary user decides whether to send a status update in each time slot with the partially observable occupation status of the spectrum. In this context, the AoI minimization problem was formulated as a POMDP. The optimal policy with threshold structure was sought by dynamic programming (DP). In [36], the authors formulated the AoI optimization problem of a status update system with a partially observable Gilbert-Elliott Channel as a belief-MDP. The authors developed an efficient structure-aware algorithm that is shown to be near-optimal. Sert et al. [37] investigated an AoI-based minimization on real-life TCP/IP connections with unknown delay and service time distributions. They trained a Deep Q-network (DQN) algorithm to perform actions on the network and obtained a near-optimal AoI performance. Reference [38] focused on the minimum-age scheduling for a time-slotted wireless uplink network, where multiple sensors are used to monitor one common physical process. The authors formulated a POMDP and analyzed the performance of a greedy policy where an AP schedules the sensor with the minimum system time in each slot. In [21] and [36], the partially observable information only have two states. Furthermore, all of the above work considered the AoI-based scheduling problem with one stream of status update. As such, the developed methods cannot be directly applied to solve our scheduling problem with multiple streams of status updates, where we need to deal with the intricate interactions of the AoI evolutions of multiple end devices. We notice that some studies have resorted to decentralized strategies for optimizing AoI in the uplink multiuser networks. By doing so, the local information at each end node can be fully observed when a node makes transmission decisions. Specifically, references [27] and [28] developed threshold-type decentralized access policies for the AoI optimization in two multiuser uplink systems. However, the decentralized policies suffer from unavoidable transmission collisions in the uplink since the transmissions of end nodes are not coordinated, which will lead to performance degradation. By contrast, this paper devises an effective centralized scheduling policy that do not need end nodes to report their status update arrivals.

We remark that part of the results presented in this work has been published in the conference version [29]. In [29], we formulated the considered scheduling design problem as a POMDP and solved it by directly applying the classical DP method. A low-complexity myopic policy was also proposed. However, the complication of the problem in its default form stopped us from conducting any theoretical analysis. In this work, we reformulate the POMDP into a belief-MDP and put forth an effective simplification of the belief-MDP. Such simplification substantially facilitate the design of the POMW policy as well as the theoretical analysis of its performance.

Notations: In this paper, $\mathbb{Z}^+$ denotes the set of non-negative integers, $\mathbb{E}[\cdot]$ denotes the operator of expectation, $[\cdot]$ denotes the representation of a vector containing the same type of elements, $\langle\cdot\rangle$ denotes a tuple containing different types of elements, and $\|\cdot\|_1$ denotes the l_1-norm of a vector.

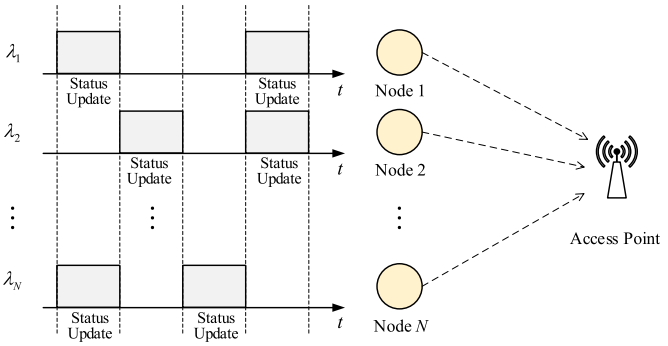


Fig. 1. The multiuser uplink system with stochastic arrival of status updates.

For two vectors, $\boldsymbol{v} = [v_l]_{l=1}^L$ and $\boldsymbol{w} = [w_l]_{l=1}^L$, with the same dimension L, $\boldsymbol{v} \geq \boldsymbol{w}$ represents $v_l \geq w_l, \forall l$.

II. System Model and POMDP Formulation

A. System Model

As shown in Fig. 1, we consider a multiuser wireless uplink network consisting of one access point (AP) and N status-updating end devices. Those end devices are also called nodes hereafter, and indexed by $i \in \{1, \ldots, N\}$. The considered system is time-slotted, and the time slot is indexed by $t \in \mathbb{Z}^+$. We consider a stochastic arrival model for the status update packets at each node. Specifically, the status update arrival at node i in each slot follows an independent and identically distributed (i.i.d.) Bernoulli process with an arrival rate λ_i. Each node maintains a single buffer to store the latest status update. That is, the current status update in the buffer will be replaced once a new one arrives. Such a single-buffer configuration, equivalent to the last-come-first-served (LCFS) queuing model, has been shown to achieve the best information freshness performance in stochastic arrival models [11], [28]. All nodes share a common wireless channel, and their transmissions of the status update packets in the uplink are coordinated by the AP. Specifically, at the beginning of each slot, the AP grants one node to transmit its latest status update packet. We denote the scheduling indicator for node i in slot t by $a_{t,i} \in \{0, 1\}$, which is equal to 1 when node i is scheduled to transmit in slot t, and $a_{t,i} = 0$ otherwise. Only one node is scheduled to transmit in each slot, thus the transmission collision among nodes is avoided. The transmission of each status update packet takes one time slot. We further assume that the transmission from node i to the AP is error-prone with a time-invariant successful rate[1] p_i.

B. Information Freshness Metric

We adopt the AoI metric, originally proposed in [10], to quantify the information freshness of all nodes at the AP. To characterize the AoI mathematically, we first define the local age $d_{t,i}$, which measures the system time of the last arrived status update packet at node i in slot t. If there is no arrival of status update at node i in the current slot, the local

[1]The considered packet loss model can implicitly incorporate practical constraints on bandwidth and peak power when we interpret the loss probability as the outage probability at physical layer.

age of the i-th node will increase by 1 at the beginning of next slot. Otherwise, the packet stored at the node is replaced by the newly arrived one, and its local age is reset to 1 at the beginning of next slot. Therefore, the evolution of $d_{t,i}$ is given by

$$d_{t+1,i} = \begin{cases} 1, & \text{if status update arrives at node } i \\ & \text{in slot } t, \\ d_{t,i}+1, & \text{otherwise.} \end{cases} \tag{1}$$

If node i is scheduled to transmit at the beginning of slot t and its transmission is successful, the local age of node i will be observed by the AP. As such, the destination AoI of node i, denoted by $D_{t,i}$, will be set to $d_{t,i}+1$ at the beginning of the next slot. Otherwise, if node i is not scheduled or the transmission fails, $D_{t,i}$ will increase by 1 at the beginning of the next slot. Mathematically, the evolution of $D_{t,i}$ is given by

$$D_{t+1,i} = \begin{cases} d_{t,i}+1, & \text{if the status update of node } i \text{ is} \\ & \text{successfully received by the AP} \\ & \text{in slot } t, \\ D_{t,i}+1, & \text{otherwise.} \end{cases} \tag{2}$$

In this paper, we assume that the local age and the destination AoI of each node are initialized as 1, i.e., $d_{0,i} = D_{0,i} = 1, \forall i$.

We remark that the local age $d_{t,i}$ and the destination AoI $D_{t,i}$ evolve independently across nodes. We consider that the AP does not grasp the specific evolutions of the local ages at all nodes and it only has the statistical arrival information (i.e., the values of λ_i's). Otherwise, the nodes need to notify each of their status update arrivals to the AP, which will lead to considerable network overhead, especially when the status update packets are relatively short. In this context, the AP only has an observation of the local age of a particular node once the node is scheduled and the transmission succeeds. Nevertheless, the AP can track the destination AoI values of all nodes, no matter whether they are scheduled or not. Overall, the AP has full information of the AoI $D_{t,i}$'s and partial observations of the local age $d_{t,i}$'s when making scheduling decisions.

C. POMDP Formulation

In this work, we adopt the long-term expected weighted sum AoI (EWSAoI) as the performance metric, which is mathematically defined as

$$\lim_{T\to\infty} \frac{1}{NT}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_i D_{t,i}\Big|\pi\right], \tag{3}$$

where $\omega_i \in (0,\infty)$ denotes the weight coefficient[2] of node i, the expectation is taken over all system dynamics, and π denotes a given multiuser scheduling policy. We aim to devise a scheduling policy π for the AP to minimize the long-term EWSAoI while fulfilling the scheduling constraint. Mathematically, we have the following optimization problem

$$\begin{aligned} \min_{\pi} \quad & \lim_{T\to\infty} \frac{1}{NT}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_i D_{t,i}\Big|\pi\right], \\ \text{s.t.,} \quad & \sum_{i=1}^{N} a_{t,i} \le 1, \quad \forall t, \end{aligned} \tag{4}$$

where the scheduling constraint is that the AP can schedule at most one node in each slot. In our design, the AP makes the scheduling decision at the beginning of each time slot. *The information available at the AP* for decision making includes the values of λ_i's, p_i's, ω_i's, the full observations of the destination AoI $D_{t,i}$'s, and the partial observations of the local age $d_{t,i}$'s. Such a decision-making problem with partial observations is naturally formulated as a POMDP. The components of the POMDP of the current model are given as follows:

- *States.* The state of node i in slot t is denoted by $\boldsymbol{s}_{t,i} \triangleq \langle d_{t,i}, D_{t,i}\rangle$, where $d_{t,i}, D_{t,i} \in \mathbb{Z}^+$. Then, the network-wide state in slot t is denoted by $\boldsymbol{s}_t \triangleq \langle \boldsymbol{d}_t, \boldsymbol{D}_t\rangle$, where $\boldsymbol{d}_t \triangleq [d_{t,1}, d_{t,2}, \ldots, d_{t,N}] \in \boldsymbol{\mathcal{D}} \triangleq (\mathbb{Z}^+)^N$ and $\boldsymbol{D}_t \triangleq [D_{t,1}, D_{t,2}, \ldots, D_{t,N}] \in \boldsymbol{\mathcal{D}}$, respectively. In addition, we denote the spaces of $\boldsymbol{s}_{t,i}$ and $\boldsymbol{s}_t$ by $\boldsymbol{\mathcal{S}}_i \triangleq \{\boldsymbol{s}_{t,i} | D_{t,i} \ge d_{t,i}\}$ and $\boldsymbol{\mathcal{S}} \triangleq \{\boldsymbol{s}_t | \boldsymbol{D}_t \ge \boldsymbol{d}_t\}$, respectively.
- *Actions.* The network-wide action in slot t is denoted by $\boldsymbol{a}_t \triangleq [a_{t,1}, a_{t,2}, \ldots, a_{t,N}]$. Recall that AP schedules at most one node in each slot, hence we have $|\boldsymbol{a}_t| \le 1$. Denote by $\boldsymbol{\mathcal{A}}$ the space of all actions, we have $\boldsymbol{a}_t \in \boldsymbol{\mathcal{A}}$.
- *Observations.* We denote the network-wide observation of the state of the nodes by $\boldsymbol{o}_t \triangleq [\boldsymbol{o}_{t,1}, \boldsymbol{o}_{t,2}, \ldots, \boldsymbol{o}_{t,N}] \in \boldsymbol{\mathcal{O}}$, where $\boldsymbol{\mathcal{O}}$ is the space of all observations. Specifically, $\boldsymbol{o}_{t,i} \triangleq \langle D_{t,i}, \hat{d}_{t,i}\rangle$ is the observation of node i in slot t, consisting of the full-observed destination AoI, $D_{t,i}$, and the partial-observed local age $\hat{d}_{t,i}$. We have $\hat{d}_{t,i} \in \mathbb{Z}^+ \bigcup \{X\}$, where X denotes no observation of the local age of node i when the node is not scheduled or the node is scheduled but the transmission fails. With these new notations, $\boldsymbol{o}_t$ can be denoted by $\langle \boldsymbol{D}_t, \hat{\boldsymbol{d}}_t\rangle$, where $\hat{\boldsymbol{d}}_t = [\hat{d}_{t,1}, \ldots, \hat{d}_{t,N}]$.
- *Transition Function.* We define the transition probability of network-wide states as $\Pr(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t)$, which denotes the conditional probability of state $\boldsymbol{s}_{t+1}$ given state $\boldsymbol{s}_t$ and action $\boldsymbol{a}_t$. We note that the transitions of $\boldsymbol{D}_t$ and $\boldsymbol{d}_t$ are conditionally independent of each other and the transition of the local age d_t is independent of the action a_t. We then have

$$\Pr(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t) = \Pr(\boldsymbol{D}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t)\Pr(\boldsymbol{d}_{t+1}|\boldsymbol{d}_t), \tag{5}$$

where

$$\Pr(\boldsymbol{D}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t) = \prod_{i=1}^{N}\Pr(D_{t+1,i}|\boldsymbol{s}_{t,i}, a_{t,i}), \tag{6}$$

and

$$\Pr(\boldsymbol{d}_{t+1}|\boldsymbol{d}_t) = \prod_{i=1}^{N}\Pr(d_{t+1,i}|d_{t,i}). \tag{7}$$

[2]The weight coefficient of a node represents the priority of status updates associated with the node.

We can further express each term on the right-hand side of (6) as

$$\Pr\left(D_{t+1,i}|\boldsymbol{s}_{t,i},a_{t,i}\right) = \begin{cases} p_i, & \text{if } a_{t,i}=1, \text{and } D_{t+1,i}=d_{t,i}+1, \\ 1-p_i, & \text{if } a_{t,i}=1, \text{and } D_{t+1,i}=D_{t,i}+1, \\ 1, & \text{if } a_{t,i}=0, \text{and } D_{t+1,i}=D_{t,i}+1, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Similarly, for each term on the right-hand side of (7), we have

$$\Pr\left(d_{t+1,i}|d_{t,i}\right) = \begin{cases} \lambda_i, & \text{if } d_{t+1,i}=1, \\ 1-\lambda_i, & \text{if } d_{t+1,i}=d_{t,i}+1, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

- *Observation Function.* Denote by $\Pr\left(\boldsymbol{o}_t|\boldsymbol{s}_t,\boldsymbol{a}_t\right)$ the network-wide observation function, which is defined as the probability of observation $\boldsymbol{o}_t$ conditioned on state $\boldsymbol{s}_t$ and action $\boldsymbol{a}_t$. Note that $\boldsymbol{D}_t$ is fully observable at the AP and the evolution of $\hat{d}_{t,i}$ with different i are independent from each other. We thus have

$$\Pr(\boldsymbol{o}_t|\boldsymbol{s}_t,\boldsymbol{a}_t)=\Pr(\hat{\boldsymbol{d}}_t|\boldsymbol{d}_t,\boldsymbol{a}_t)=\prod_{i=1}^{N}\Pr(\hat{d}_{t,i}|d_{t,i},a_{t,i}), \tag{10}$$

where we term

$$\Pr\left(\hat{d}_{t,i}|d_{t,i},a_{t,i}\right)=\begin{cases} p_i, & \text{if } \hat{d}_{t,i}=d_t \text{ and } a_{t,i}=1, \\ 1-p_i, & \text{if } \hat{d}_{t,i}=X \text{ and } a_{t,i}=1, \\ 1, & \text{if } \hat{d}_{t,i}=X \text{ and } a_{t,i}=0, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

as the local age observation function of node i.
- *Immediate Reward.* We target to optimize the long-term EWSAoI. Based on that, We define the immediate reward of state $\boldsymbol{s}_t$ as $r\left(\boldsymbol{s}_t\right) \triangleq \sum_{i=1}^{N}\omega_i D_{t,i}$.

We remark that due to the partially observed network-wide state $\boldsymbol{s}_t$, the formulated POMDP problem cannot be solved by directly applying the existing AoI-oriented scheduling frameworks designed for the scenarios with full observation of network-wide states (e.g., [22], [26], [28]). To circumvent the problem, we will leverage the sufficient posterior probability distribution of $\boldsymbol{s}_t$ with the observation $\boldsymbol{o}_t$ at the AP. Such probability distributions are also named as the belief states of the POMDP [35]. In the following, we will reformulate our POMDP as a belief-MDP, where the belief states of the POMDP are regarded as the states of the belief-MDP.

III. Belief-MDP Formulation and Simplification

In this section, we first reformulate the POMDP introduced in Section II as a belief-MDP and then simplify the belief-MDP to gain more insights.

A. Reformulation of the POMDP

With reference to [39], a POMDP can be converted to an equivalent belief-MDP based on the belief states of the system. To that end, we now introduce the definitions of the belief states and other components of the belief-MDP version of our POMDP problem as follows:

- *Belief States.* The belief state of node i is defined as the current probability distribution over $\boldsymbol{\mathcal{S}}_i$ on condition of the history so far. Mathematically, the belief state of node i in slot t is denoted by

$$\boldsymbol{B}_{t,i} \triangleq \left[B_{t,i}(\boldsymbol{s}_{t,i})\right]_{\boldsymbol{s}_{t,i}\in\boldsymbol{\mathcal{S}}_i} \tag{12}$$

with $\|\boldsymbol{B}_{t,i}\|_1$=1, where $B_{t,i}(\boldsymbol{s}_{t,i}) \triangleq \Pr\left(\boldsymbol{s}_{t,i}|\boldsymbol{h}_{t,i}\right)$ denotes the probability assigned to state $\boldsymbol{s}_{t,i}$ with the current history $\boldsymbol{h}_{t,i} \triangleq \langle \boldsymbol{B}_{1,i}, a_{1,i}, \boldsymbol{o}_{1,i}, a_{2,i}, \ldots, a_{t-1,i}, \boldsymbol{o}_{t-1,i}\rangle$ of node i. As mentioned in Section II-B, $D_{t,i}$ is deterministic for a given history profile $\boldsymbol{h}_{t,i}$ since $\boldsymbol{h}_{t,i}$ includes $\boldsymbol{o}_{t-1,i}$. Therefore, $\boldsymbol{B}_{t,i}$ can also be represented by $\langle D_{t,i}, \boldsymbol{b}_{t,i}\rangle$, where $\boldsymbol{b}_{t,i} \triangleq \left[b_{t,i}(d_{t,i})\right]_{d_{t,i}\in\mathbb{Z}^+}$ denotes the belief state of the local age of node i, and $\|\boldsymbol{b}_{t,i}\|_1 = 1$. Furthermore, $b_{t,i}\left(d_{t,i}\right) \triangleq \Pr\left(d_{t,i}|\boldsymbol{h}_{t,i}\right)$ denotes the probability assigned to $d_{t,i}$. Hence, we have $B_{t,i}\left(\boldsymbol{s}_{t,i}\right) = b_{t,i}\left(d_{t,i}\right)$ given $D_{t,i}$.
The network-wide belief state is defined as the current probability distribution over $\boldsymbol{\mathcal{S}}$ on condition of $\boldsymbol{h}_t \triangleq \langle \boldsymbol{B}_1, a_1, \boldsymbol{o}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{t-1}, \boldsymbol{o}_{t-1}\rangle$, and it is also the state of the belief-MDP. We denote the network-wide belief state in slot t by

$$\boldsymbol{B}_t \triangleq \left[B_t(\boldsymbol{s}_t)\right]_{\boldsymbol{s}_t\in\boldsymbol{\mathcal{S}}} = \langle \boldsymbol{D}_t, \boldsymbol{b}_t\rangle, \tag{13}$$

where $\boldsymbol{b}_t \triangleq \left[b_t(\boldsymbol{d}_t)\right]_{\boldsymbol{d}_t\in\boldsymbol{\mathcal{D}}}$ is the belief state of all local ages in slot t with $b_t\left(\boldsymbol{d}_t\right) \triangleq \Pr\left(\boldsymbol{d}_t|\boldsymbol{h}_t\right)$ denoting the probability[3] assigned to $\boldsymbol{d}_t$, and with $B_t(\boldsymbol{s}_t) \triangleq \Pr\left(\boldsymbol{s}_t|\boldsymbol{h}_t\right)$ denoting the probability assigned to $\boldsymbol{s}_t$. Thus, we have $\|\boldsymbol{B}_t\|_1 = \|\boldsymbol{b}_t\|_1 = 1$. With a given $\boldsymbol{D}_t$, the belief state of the local age of each node evolves independently in our POMDP framework, and thus we have $B_t\left(\boldsymbol{s}_t\right) = b_t\left(\boldsymbol{d}_t\right) = \prod_{i=1}^{N} b_{t,i}\left(d_{t,i}\right)$. Besides, we denote $\boldsymbol{\mathcal{B}}$ as the belief space, i.e., the collection of all possible $\boldsymbol{B}_t$. $\boldsymbol{\mathcal{B}}$ is also called *belief simplex* [40].
- *Belief Update.* The AP can update $\boldsymbol{B}_{t+1}$ from $\boldsymbol{B}_t$ at the end of slot t after receiving new observations once the last action $\boldsymbol{a}_t$ is executed. Recall that $\boldsymbol{B}_t = \langle \boldsymbol{D}_t, \boldsymbol{b}_t\rangle$, both $\boldsymbol{D}_t$ and $\boldsymbol{b}_t$ need to be updated. Specifically, the destination AoI of node i, i.e., the i-th component of $\boldsymbol{D}_t$, can be updated by

$$D_{t+1,i} = \begin{cases} D_{t,i}+1, & \text{if } \hat{d}_{t,i}=X, \\ \hat{d}_{t,i}+1, & \text{otherwise.} \end{cases} \tag{14}$$

The update of $D_{t,i}$ is deterministic and independent from node to node. Moreover, $\boldsymbol{b}_{t+1}$ can be updated from $\boldsymbol{b}_t$ through the Bayes' theorem as

$$b_{t+1}(\boldsymbol{d}_{t+1}) = \rho \sum_{\boldsymbol{d}_t\in\boldsymbol{\mathcal{D}}} b_t\left(\boldsymbol{d}_t\right)\Pr\left(\boldsymbol{d}_{t+1}|\boldsymbol{d}_t\right)\Pr\left(\hat{\boldsymbol{d}}_t|\boldsymbol{d}_t,\boldsymbol{a}_t\right), \tag{15}$$

[3]We omitted $\boldsymbol{h}_{t,i}$ in the definition of the belief state for concise notation.

where

$$\rho = 1/\sum_{\boldsymbol{d}_{t+1},\boldsymbol{d}_t\in\boldsymbol{\mathcal{D}}} b_t\left(\boldsymbol{d}_t\right)\Pr\left(\boldsymbol{d}_{t+1}|\boldsymbol{d}_t\right)\Pr\left(\hat{\boldsymbol{d}}_t|\boldsymbol{d}_t,\boldsymbol{a}_t\right) \tag{16}$$

is the Bayes normalizing factor. Considering the independent evolutions of $d_{t,i}$'s across nodes, we can also update $\boldsymbol{b}_t$ via updating $\boldsymbol{b}_{t,i}$ of each node i individually. We omit the update equation of $\boldsymbol{b}_{t,i}$ here for brevity.

- *Actions.* The action of the belief-MDP in slot t is denoted by $\boldsymbol{a}_t \in \boldsymbol{\mathcal{A}}$, which is exactly same as that of the POMDP.
- *Transition Function.* The transition function of the belief-MDP is given by

$$\Pr\left(\boldsymbol{B}_{t+1}|\boldsymbol{B}_t,\boldsymbol{a}_t\right) = \sum_{\boldsymbol{o}_t\in\boldsymbol{\mathcal{O}}}\Pr\left(\boldsymbol{B}_{t+1}|\boldsymbol{B}_t,\boldsymbol{a}_t,\boldsymbol{o}_t\right)\Pr\left(\boldsymbol{o}_t|\boldsymbol{B}_t,\boldsymbol{a}_t\right), \tag{17}$$

where

$$\Pr(\boldsymbol{o}_t|\boldsymbol{B}_t,\boldsymbol{a}_t) = \sum_{\boldsymbol{s}_t\in\boldsymbol{\mathcal{S}}} B_t(\boldsymbol{s}_t)\Pr(\boldsymbol{o}_t|\boldsymbol{s}_t,\boldsymbol{a}_t), \tag{18}$$

and

$$\Pr\left(\boldsymbol{B}_{t+1}|\boldsymbol{B}_t,\boldsymbol{a}_t,\boldsymbol{o}_t\right) = \begin{cases} 1, & \text{if the belief update with arguments} \\ & \boldsymbol{B}_t,\boldsymbol{a}_t,\boldsymbol{o}_t \text{ returns } \boldsymbol{B}_{t+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

- *Policy.* We adopt a deterministic stationary scheduling policy π for the belief-MDP. The policy maps the belief space $\boldsymbol{\mathcal{B}}$ to the action space in each slot.
- *Reward.* Since the destination AoI is deterministic for the AP, the immediate expected reward on condition of belief state $\boldsymbol{B}_t$ is the same as that in the POMDP, i.e., $R\left(\boldsymbol{B}_t\right) \triangleq \mathbb{E}\left[r(\boldsymbol{s}_t)|\boldsymbol{B}_t\right] = \sum_{i=1}^{N}\omega_i D_{t,i}$. On this basis, the objective problem can be rewritten as

$$\min_{\pi}\ \lim_{T\to\infty}\frac{1}{NT}\mathbb{E}\left[\sum_{t=1}^{T} R(\boldsymbol{B}_t)\Big|\boldsymbol{B}_1,\pi\right], \quad \text{s.t., } \|\boldsymbol{a}_t\|_1 \le 1, \quad \forall t, \tag{20}$$

where $\boldsymbol{B}_1$ is a predefined initial belief state. Recall that we assume $d_{0,i} = D_{0,i} = 1, \forall i$, before running the network, and thus $\boldsymbol{B}_{1,i} = \langle 2, [\lambda_i, 1-\lambda_i, 0, \cdots]\rangle, \forall i$.

We remark that the belief update is computationally complicated when the dimension of the belief states is high, and is impractical when the dimension goes to infinity. Moreover, the continuousness of the belief space $\boldsymbol{\mathcal{B}}$ leads to a PSPACE hardness of optimizing the EWSAoI of the belief-MDP optimally [35]. We have explored the dynamic programming (DP) approach to optimally solve the formulated POMDP with finite horizons in the conference version of this work [29]. Our findings in [29] showed that the optimal policy can only be used for a system with a very small number of users due to the high computational complexity. This motivates us to further simplify the belief-MDP by exploring its structures. For brevity, we omit the details of the method proposed in [29].

B. Belief-MDP Simplification

We subsequently show the existence of a simplified representation of the belief-MDP with the given $\boldsymbol{B}_1$. To start, we have the following definition:

Definition 1: Assume AP schedules node i in slot t with observation $d_{t,i} = k_i$, and then does not receive any packet from node i in the following m_i slots. Define the local age belief state of node i in slot $t + m_i$ by $\boldsymbol{c}(k_i, m_i)$, namely, the belief of node i with the last observation k_i followed by m_i elapsed slots.

For convenience, we ignore index i for nodes and introduce the following proposition.

Proposition 1: The distribution vector of the local age belief state $\boldsymbol{c}(k,m)$ of node i in slot t can be given by

$$\begin{aligned}\boldsymbol{c}(k,m) &= \left[c_{k,m}\left(d_t\right)\right]_{d_t\in\mathbb{Z}^+} \\ &= \left[\lambda, \lambda\gamma, \lambda\gamma^2, \cdots, \lambda\gamma^{m-1}, 0, \cdots, 0, \gamma^m, 0, \cdots\right],\end{aligned} \tag{21}$$

where $k, m \in \mathbb{Z}^+$, $\gamma = 1-\lambda$, and $c_{k,m}\left(d_t\right)$ denotes the probability assigned to d_t. The position of entry γ^m is $k+m$, and this denotes that the corresponding destination AoI of entry γ^m is $k+m$.

Proof: See Appendix A of [41]. □

We remark that **Proposition 1** shares the same spirit as the characterization of the local age belief states given in [38, **Proposition** 4]. However, reference [38] considered the evolution of the local ages of one stream. In contrast, we consider the evolutions of the destination AoI and the local ages of multiple streams in our system. The strategy in [38] cannot coordinate the AoI of different streams in our model, and thus it cannot be used to solve the formulated POMDP. Moreover, a truncation was applied to the local ages in [38] but is not used in this paper. On the other hand, our idea of the belief state simplification is similar to that in [42]. However, the detailed expressions of the simplified belief states are not provided in [42], and assumptions on the state transition made in [42] do not hold in our case. As such, the method in [42] can also not be applied to solve our POMDP.

Define a group of belief states that have the AoI equal to $k+m$ together with $\boldsymbol{c}(k,m)$ defined in **Proposition 1** as $\boldsymbol{C}(k,m) \triangleq \langle k+m, \boldsymbol{c}(k,m)\rangle$ for $m, k \in \mathbb{Z}^+$. Denote by $\boldsymbol{\mathcal{C}}$ the collection of all possible $\boldsymbol{C}(k,m)$. Then, we have the following corollary.

Corollary 1: Suppose the network has a certain belief state, i.e., $\boldsymbol{b}_{0,i} = \boldsymbol{e}_1, D_{0,i} = 1, \forall i$ before running, then $\boldsymbol{B}_{t,i} \in \boldsymbol{\mathcal{C}}$ for $t = 1, 2, \cdots, T, \forall i$.

Proof: We use induction to prove it. First, it is clear that $\boldsymbol{b}_{1,i} = \boldsymbol{c}(1,1), D_{1,i} = 2$, and hence $\boldsymbol{B}_{1,i} \in \boldsymbol{\mathcal{C}}$. In slot t, suppose $\boldsymbol{B}_{t,i} = \langle k_{t,i} + m_{t,i}, \boldsymbol{c}(k_{t,i}, m_{t,i})\rangle \in \boldsymbol{\mathcal{C}}, \forall i$, where $k_{t,i}$ and $m_{t,i}$ denote the last observation of local age and the number of slots elapsed since the last observation of node i in slot t, respectively. Now, we consider following cases:

- The AP schedules node i, and the transmission of node i succeeds. In this case, the AP observes the local age

of node i. By (21), only entries $1, 2, \cdots, m_{t,i}$ and $k_{t,i} + m_{t,i}$ in $\boldsymbol{b}_{t,i}$ are greater than 0, and hence we have all possible observations of the local age given by $\hat{k}_{t,i} \in \{1, 2, \cdots, m_{t,i}\} \cup \{k_{t,i} + m_{t,i}\}$. And by (14), we have $D_{t+1,i} = \hat{k}_{t,i} + 1$. Hence, we have $\boldsymbol{B}_{t+1,i} = [\hat{k}_{t,i} + 1, \boldsymbol{c}(\hat{k}_{t,i}, 1)] \in \boldsymbol{\mathcal{C}}$.

- The AP schedules node i, and the transmission of node i fails. In this context, the AP cannot observe the local age of node i in slot t, and hence the elapsed time of no observation will become $m_{t,i} + 1$ in slot $t + 1$, and $D_{t+1,i} = D_{t,i} + 1 = k_{t,i} + m_{t,i} + 1$. Hence, we have $\boldsymbol{B}_{t+1,i} = [k_{t,i} + m_{t,i} + 1, \boldsymbol{c}(k_{t,i}, m_{t,i} + 1)] \in \boldsymbol{\mathcal{C}}$.
- For any node that is not scheduled, its belief state will be still in class $\boldsymbol{\mathcal{C}}$ in the next time slot. This is because the transition of this node is same as that of a scheduled node whose transmission fails.

The above three cases cover all possible results in slot $t + 1$. Hence, we have $\boldsymbol{B}_{t+1,i} \in \boldsymbol{\mathcal{C}}, \forall i$. This completes the proof. □

Based on ***Corollary* 1**, each infinite dimensional belief state $\boldsymbol{B}_{t,i} \in \boldsymbol{\mathcal{C}}$ can be sufficiently represented by two positive integers $k_{t,i}$ and $m_{t,i}$, with $D_{t,i} = k_{t,i} + m_{t,i}$ and $\boldsymbol{b}_{t,i} = \boldsymbol{c}(k_{t,i}, m_{t,i})$. Hence, the belief-MDP framework in **Section III** can be characterized in a much simpler form. We name this simplified representation of belief MDP as Last-Observation-Characterized (LOC) belief-MDP. The actions of the LOC belief-MDP are the same as that of the original belief-MDP. The other components of the LOC belief-MDP are presented as follows.

- *States.* The state of node i in slot t is denoted by $\boldsymbol{z}_{t,i} \triangleq [k_{t,i}, m_{t,i}]$, where $k_{t,i}, m_{t,i} \in \mathbb{Z}^+$ are defined in ***Corollary* 1**. Then, the network-wide state in slot t is denoted by $\boldsymbol{z}_t \triangleq [\boldsymbol{k}_t, \boldsymbol{m}_t]$, where $\boldsymbol{k}_t \triangleq [k_{t,1}, k_{t,2}, \ldots, k_{t,N}] \in \boldsymbol{\mathcal{D}}$ and $\boldsymbol{m}_t \triangleq [m_{t,1}, m_{t,2}, \ldots, m_{t,N}] \in \boldsymbol{\mathcal{D}}$. Define $\boldsymbol{\mathcal{Z}} \triangleq \boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{D}}$ as the space set of $\boldsymbol{z}_t$. $\boldsymbol{\mathcal{Z}}$ also corresponds to the feasible part of the belief space $\boldsymbol{\mathcal{B}}$ for belief states with the initialization in ***Corollary* 1**.
- *Transition Function.* We define the transition function of the LOC belief-MDP as $\Pr(\boldsymbol{z}_{t+1}|\boldsymbol{z}_t, \boldsymbol{a}_t)$, which is given by

$$\Pr(\boldsymbol{z}_{t+1}|\boldsymbol{z}_t, \boldsymbol{a}_t) = \prod_{i=1}^{N} \Pr(\boldsymbol{z}_{t+1,i}|\boldsymbol{z}_{t,i}, a_{t,i}), \tag{22}$$

where

$$\Pr(\boldsymbol{z}_{t+1,i}|\boldsymbol{z}_{t,i}, a_{t,i} = 1) = \begin{cases} p_i\lambda_i(1-\lambda_i)^{d-1}, & \text{if } k_{t+1,i} = d \text{ and } m_{t+1,i} = 1, \\ p_i(1-\lambda_i)^{m_{t,i}}, & \text{if } k_{t+1,i} = k_{t,i} + m_{t,i} \text{ and } m_{t+1,i} = 1, \\ 1 - p_i, & \text{if } k_{t+1,i} = k_{t,i} \text{ and } m_{t+1,i} = m_{t,i} + 1, \\ 0, & \text{otherwise}, \end{cases} \tag{23}$$

with $d \in \{1, 2, \cdots, m_{t,i}\}$. Furthermore, $\Pr(\boldsymbol{z}_{t+1,i}|\boldsymbol{z}_{t,i}, a_{t,i} = 0) = 1$ if $k_{t+1,i} = k_{t,i}$ and $m_{t+1,i} = m_{t,i}+1$.

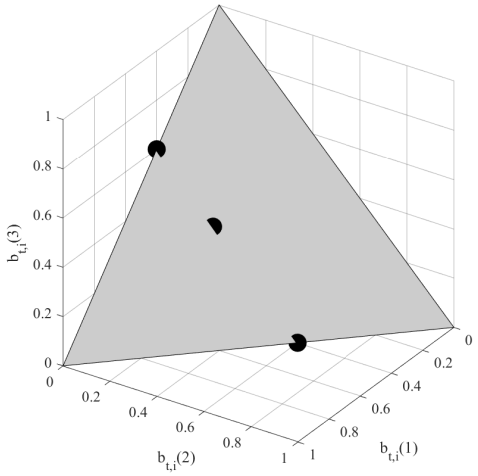


Fig. 2. The sub-region of $\boldsymbol{\mathcal{B}}$ and its reduced feasible space, i.e., points $[0.4, 0.6, 0, \cdots]$,$[0.4, 0, 0.6, 0, \cdots]$, and $[0.4, 0.24, 0.36, 0, \cdots]$ in the three-dimensional space with $N = 1$ and $\lambda = 0.4$.

- *Reward.* The expected immediate reward given a state $\boldsymbol{z}_t$ is rewritten as $R(\boldsymbol{z}_t) \triangleq \sum_{i=1}^{N} \omega_i(k_{t,i} + m_{t,i})$.
- *Policy.* The policy for the LOC belief-MDP framework is the same as that in **Section III** with a different domain $\boldsymbol{\mathcal{Z}}$. It can be equivalently denoted by $\pi : \boldsymbol{\mathcal{Z}} \longmapsto \boldsymbol{\mathcal{A}}$.

In typical work on solving a belief-MDP, one need to use the *Backup operation* [40], [43] to repeatedly find more feasible belief states and update the feasible belief space horizon by horizon. It is computationally complicated, and unlikely to reach most of feasible belief states in the belief simplex. However, with the above simplification, we reduce the space of belief states sharply from the continuous space $\boldsymbol{\mathcal{B}}$ to a discrete space $\boldsymbol{\mathcal{Z}}$. That enables us to directly obtain the full feasible space of the belief states without using the inefficient *Backup operation*. Furthermore, the completed transition probabilities of belief states can be obtained by (22). Fig.2 illustrates one example of the space reduction, where we have one node with its status update arrival rate $\lambda = 0.4$. The gray triangle plane is the sub-region of $\boldsymbol{\mathcal{B}}$ in the three-dimensional space, on which each point is associated with a possible local age belief state. After the simplification, the sub-region of the belief space $\boldsymbol{\mathcal{B}}$ can be reduced to three feasible belief states, i.e., the three points plotted on the sub-region.

Remark 1: We can extend the above LOC belief-MDP simplification process to the scenario with Markovian arrival processes. Specifically, the belief states of a node can still be characterized by two-dimensional vectors. More details can be found in Appendix D of [41].

IV. POMW Policy

Based on the LOC belief-MDP, we can use the conventional DP approach to solve the AoI scheduling problem. However, the LOC belief-MDP is formulated for a multiuser model, thus the DP would suffer from the curse of the dimensionality as the number of end devices increases. To circumvent such a problem, we propose a low-complexity policy for the EWSAoI optimization in the considered network with partial observations, named POMW policy.

We remark that a downlink network with the same status update traffic model as ours was investigated in [28]. Different from our network, the local age of the status update packets are fully observable at the AP due to the downlink setting. The authors devised an Age-based Max-Weight policy by leveraging the Lyapunov Optimization [34]. This policy minimizes a defined Lyapunov drift on condition of the fully observable local age and destination AoI in each slot. Hereafter, we call it Fully Observable Max-Weight (FOMW) policy. Moreover, for brevity, we use *"FON"* to represent the network with the fully observable states in [28] and *"PON"* to represent our considered network with partial observations in the rest of this paper.

Inspired by [28], we apply the Lyapunov Optimization to devise a low-complexity policy, i.e., the POMW policy, which extends the FOMW policy developed in [28]. To that end, we will define a Lyapunov Function based on the EWSAoI of the network. The POMW policy attempts to minimize the expected drift of the Lyapunov Function under condition of the current belief state and destination AoI in each slot t. Therefore, the EWSAoI of the network can be optimized with lower computational complexity.

We define the linear Lyapunov Function as

$$L(t) = \frac{1}{N}\sum_{i=1}^{N} \beta_i D_{t,i}, \tag{24}$$

where $\beta_i > 0$ is an hyper-parameter that can be used to tune the POMW policy to different network configurations. The Lyapunov Drift is defined as

$$\Delta\left[\boldsymbol{B}_t\right] = \mathbb{E}\left[L(t+1) - L(t)|\boldsymbol{B}_t\right]. \tag{25}$$

The Lyapunov Drift $\Delta\left[\boldsymbol{B}_t\right]$ refers to the expected increase of the Lyapunov Function $L(t)$ in one slot. Hence, by minimizing the drift in (25), the POMW policy equivalently reduces $L(t)$. Consequently, the EWSAoI of the network is kept low.

To develop the POMW policy for the Lyapunov Drift minimization, we analyze the expression for the drift in (25). Recall the definition of $\boldsymbol{B}_t$, we realize that the value of $L(t)$ is fixed with a given $\boldsymbol{B}_t$. Thus minimizing the Lyapunov Drift in (25) is equivalent to minimizing $\mathbb{E}[L(t+1)|\boldsymbol{B}_t]$. Recall the evolution of $D_{t,i}$ given in (2), and we have

$$\begin{aligned}
&\mathbb{E}[L(t+1)|\boldsymbol{B}_t] \\
&= \sum_{i=1}^{N} \frac{1}{N}\mathbb{E}\left[\beta_i D_{t+1,i}\middle|\boldsymbol{B}_t\right] \\
&= \sum_{i=1}^{N} \frac{\beta_i}{N}\left[p_i a_{t,i}\sum_{d\in\mathbb{Z}^+} b_{t,i}(d)(d+1) + (1-p_i a_{t,i})(D_{t,i}+1)\right] \\
&= \frac{1}{N}\left[-\sum_{i=1}^{N} a_{t,i}\beta_i p_i G_{t,i} + \sum_{i=1}^{N}\beta_i(D_{t,i}+1)\right],
\end{aligned} \tag{26}$$

where

$$G_{t,i} := D_{t,i} - \sum_{d\in\mathbb{Z}^+} b_{t,i}(d)d. \tag{27}$$

Eq. (26) leads to following proposition:

Proposition 2: To minimize the Lyapunov Drift in slot t, the POMW policy should schedule node i with the maximal $\beta_i p_i G_{t,i}$.

The proof of **Proposition 2** is straightforward and hence is omitted. Before the POMW policy making the scheduling decision, the belief probabilities $b_{t,i}(d)$ need to be updated based on the observations of the previous slot.

Remark 2: Note that when local age is fully observed, we have

$$\sum_{d\in\mathbb{Z}^+} b_{t,i}(d)d = d_{t,i}, \forall t, i. \tag{28}$$

In this case, the POMW policy will schedule node i with the maximal $\beta_i p_i \left(D_{t,i} - d_{t,i}\right)$ in each slot, which exactly coincides with the criterion of the FOMW policy presented in [28]. This observation indicates that the POMW policy is a generalization of the FOMW policy.

However, it is hard to implement this online policy on the fly due to the high computational complexity. In each slot, the POMW policy selects an action $\boldsymbol{a}_t$ by minimizing (26). This step requires $O\left(N|\boldsymbol{\mathcal{A}}||\mathbb{Z}^+|\right)$ operations. Subsequently, the policy updates the local age belief states for the next slot by the Bayes' theorem. Such an update step requires $O\left(N|\mathbb{Z}^+|^2\right)$ operations. Those two steps are computationally intractable since $\mathbb{Z}^+$ is an infinite set[4]. Thus, the straightforward application of the FOMW policy to our problem could be impractical.

Thanks to the LOC belief-MDP framework proposed in **Proposition 1**, we are able to simplify the expression of $G_{t,i}$ from complex expectation calculation to a closed-form expression of only three parameters. More specifically, it can be expressed as

$$\begin{aligned}
G_{t,i} &= k_{t,i} + m_{t,i} - \boldsymbol{c}(k_{t,i}, m_{t,i})\boldsymbol{n} \\
&= m_{t,i} + \left[1 - (1-\lambda_i)^{m_{t,i}}\right]\left(k_{t,i} - \frac{1}{\lambda_i}\right),
\end{aligned} \tag{29}$$

where $\boldsymbol{n} = [1, 2, 3, \cdots]^T$. By now, we can formally describe the POMW policy in **Algorithm 1**. The POMW policy can minimize the Lyapunov Drift with low computational complexity, and consequently optimize the EWSAoI of the network.

Note that in **Algorithm 1**, the updates of $k_{t,i}$ and $m_{t,i}$ are based on the transition function of the LOC belief-MDP given in (23).

Thanks to the proposed simplification, the complexity of the step to select an action is reduced from $O\left(N|\boldsymbol{\mathcal{A}}||\mathbb{Z}^+|\right)$ to $O\left(N|\boldsymbol{\mathcal{A}}|\right)$. The complexity of updating states is reduced from $O\left(N|\mathbb{Z}^+|^2\right)$ to $O\left(2N\right)$. Moreover, we do not need to set a truncation on destination AoI or local age when implementing **Algorithm 1**.

V. Performance Analyses

In this section, we first introduce a low-complexity policy named Randomized Scheduling (RS) policy and analyze its

[4]One can truncate the maximum value of AoI to make the computation feasible. However, a sufficiently large cap of the AoI should be applied to ensure the accuracy of the truncation, which still leads to unacceptably high computational complexity.

Algorithm 1 POMW Policy

Initialization: $t=1, m_{t,i}=1, k_{t,i}=1, \forall i$[5];
while *each new slot* t **do**
  for *each node* i **do**
    $G_{t,i} = m_{t,i} + [1-(1-\lambda_i)^{m_{t,i}}]\left(k_{t,i} - \frac{1}{\lambda_i}\right)$;
  end
  Schedule node j in the current slot, where $j = \arg\max_i \beta_i p_i G_{t,i}$;
  Obtain the local age observation $\hat{d}_{t,j}$ of node j;
  if $\hat{d}_{t,j} = X$ **then**
    $m_{t+1,j} = m_{t,j} + 1$;
  else if $\hat{d}_{t,j} = d \in \mathbb{Z}^+$ **then**
    $m_{t+1,j} = 1, k_{t+1,j} = d$;
  end
  $m_{t+1,i} = m_{t,i} + 1, \forall i \neq j$;
  $t = t+1$;
end

EWSAoI performance. Based on its performance, we derive the upper bounds for the EWSAoI performance of the POMW policy. We also analyze the performance guarantee of the POMW policy by comparing its EWSAoI performance with a universal lower bound in the literature.

A. RS Policy

In [28], an RS policy was proposed to optimize the AoI of the network. In the RS policy, node i is scheduled with probability $\mu_i \in (0,1]$ in each slot. The scheduling probabilities are time-invariant and satisfy $\sum_{i=1}^N \mu_i \leq 1$. Notice that the actions of the RS policy is independent of the network states, thus this policy can also be adopted in the considered PON. With reference to the proof of [28, Prop. 4], we give the EWSAoI performance of the RS policy, denoted by R^{RS}, in ***Proposition* 3**.

Proposition 3: The EWSAoI of the network under the RS policy with scheduling probabilities $\{\mu_i\}_{i=1}^N$ *is*

$$R^{RS} = \frac{1}{N}\sum_{i=1}^N \omega_i \left(\frac{1}{\lambda_i} + \frac{1}{p_i\mu_i}\right). \tag{30}$$

Note that the EWSAoI in (30) is slightly different from that in [28] due to the difference in the local age evolution of two systems. Denote by $\{\mu_i^*\}_{i=1}^N$ the optimal scheduling probabilities of all node, the optimal RS policy is given as follows [28, Th. 5].

Theorem 1: Consider the network under the RS policy. The optimal scheduling probabilities are

$$\mu_i^* = \frac{\sqrt{\omega_i/p_i}}{\sum_{j=1}^N \sqrt{\omega_j/p_j}}, \tag{31}$$

[5]To ease understanding and simplify expressions, we set such an initialization. Without loss of generality, we can also select any $\boldsymbol{b}_1 \in \boldsymbol{\mathcal{B}}$ for the initialization. In that case, when $m_{t,i}$ and $k_{t,i}$ do not exist for some i,t, we can update belief states by (15) and calculate $G_{t,i}$ by (27).

and correspondingly,

$$R^{RS^*} = \frac{1}{N}\left[\sum_{i-1}^N \frac{\omega_i}{\lambda_i} + \left(\sum_{i-1}^N \sqrt{\frac{\omega_i}{p_i}}\right)^2\right]. \tag{32}$$

According to [28, Th.10], R^{RS*} is the upper bound of the EWSAoI performance of the FOMW policy, denoted by R^{FOMW}. The FOMW policy can be regarded as the full-observed counterpart of the POMW policy. Nevertheless, the analytical method in [28] cannot be directly applied to derive upper bounds for the POMW policy as it will not lead to any insightful results. The rationale is that the analysis of the POMW policy is more challenging due to the complicated transitions between the belief states in the continuous space. Thanks to the proposed LOC simplification, we manage to derive two upper bounds for the POMW policy given in the subsequent subsections.

B. Upper Bounds of the POMW Policy

Built upon the proposed LOC belief-MDP, we now derive two upper bounds for the POMW policy. One of them is the EWSAoI performance of a particular RS policy, as stated in the following theorem:

Theorem 2: The EWSAoI performance of the POMW policy with $\beta_i = \omega_i/\lambda_i\mu_i' p_i, \forall i$, *denoted by* R^{POMW}, *is upper bounded by*

$$R^{POMW} \leq \frac{1}{N}\sum_{i=1}^N \omega_i \left(\frac{1}{\lambda_i\mu_i' p_i} + 1\right) \leq R^{RSM}, \tag{33}$$

where

$$\mu_i' = \frac{\sqrt{\omega_i/\lambda_i p_i}}{\sum_{j=1}^N \sqrt{\omega_j/\lambda_j p_j}}, \forall i, \tag{34}$$

are a series of scheduling probabilities of all nodes in the network. R^{RSM} *is the EWSAoI of an RS policy with the corresponding scheduling probabilities*

$$\mu_i^M = \frac{\sqrt{\omega_i\lambda_i/p_i}}{\sum_{j=1}^N \sqrt{\omega_j/\lambda_j p_j}}, \forall i. \tag{35}$$

Proof: We prove it by leveraging the introduced RS policy. See Appendix B of [41] for details. □

Note that the value assigned to β_i, which depends on μ_i', can be attained by minimizing the upper bound of the EWSAoI, given by $\frac{1}{N}\sum_{i=1}^N \omega_i\,(1/(\lambda_i\mu_i' p_i) + 1)$. A similar method was used in [34].

C. Performance Guarantee of the POMW Policy

Based on ***Theorem* 2**, we can analyze the performance guarantee of the POMW policy theoretically. Firstly, we introduce a universal lower bound given in [28, Th.3]. This lower bound applies to any feasible scheduling policy, and is applicable to both FONs and PONs. The universal lower bound of the

TABLE I
$\boldsymbol{c}(k,m)$: SIMULATION RESULTS VERSUS THEORETICAL RESULTS WITH $\lambda = 0.6$

Belief states of the local age	$c_{k,m}(1)$	$c_{k,m}(2)$	$c_{k,m}(3)$	$c_{k,m}(4)$	$c_{k,m}(5)$
$\boldsymbol{c}(1,4)$ (simulation)	0.60322	0.2384	0.0958	0.037	0.02558
$\boldsymbol{c}(1,4)$ (theoretical)	0.6	0.24	0.096	0.0384	0.0256
$\boldsymbol{c}(2,3)$ (simulation)	0.59744	0.24156	0.09602	0	0.06498
$\boldsymbol{c}(2,3)$ (theoretical)	0.6	0.24	0.096	0	0.064
$\boldsymbol{c}(3,2)$ (simulation)	0.59832	0.24116	0	0	0.16052
$\boldsymbol{c}(3,2)$ (theoretical)	0.6	0.24	0	0	0.16
$\boldsymbol{c}(4,1)$ (simulation)	0.59756	0	0	0	0.40244
$\boldsymbol{c}(4,1)$ (theoretical)	0.6	0	0	0	0.4

EWSAoI performance of any policies is given by

$$L_B = \min_{\{q_i\}_{i=1}^N} \frac{1}{2N}\sum_{i=1}^{N}\omega_i\left(\frac{1}{q_i}+3\right), \tag{36}$$

$$\text{s.t.,} \sum_{i=1}^{N} q_i/p_i \le 1, \tag{37}$$

$$q_i \le \lambda_i, \forall i. \tag{38}$$

The solution q_i^* of the above problem can be obtained following [28, Algorithm 1], and the lower bound is

$$L_B = \frac{1}{2N}\sum_{i=1}^{N}\omega_i\left(\frac{1}{q_i^*}+3\right). \tag{39}$$

Based on the lower bound (39), we can have the following corollary on the performance guarantee of the POMW policy.

Corollary 2: The performance of the POMW policy with $\beta_i = \omega_i/\lambda_i q_i^*$ follows that

$$\frac{R^{POMW}}{L_B} < \frac{2}{\lambda_{min}}, \tag{40}$$

where $\lambda_{min} \triangleq \min\{\lambda_i\}_{i=1}^N$.

Proof: See Appendix C of [41]. □

*Remark 3: We use the ratio between R^{POMW} and L_B to evaluate the performance guarantee of the POMW policy. **Corollary 2** indicates that the ratio is inversely proportional to the packet arrival rates of nodes λ_i in the network. When the network is close to the "generate-at-will", i.e., $\lambda_i \to 1, \forall i$, the ratio of R^{POMW} and L_B with $\beta_i = \omega_i/\lambda_i q_i^*$ is smaller than 2. This coincides with the performance guarantee of a counterpart FOMW policy devised for the "generate-at-will" system in [13].*

VI. NUMERICAL RESULTS

In this section, we first provide some numerical results on belief states defined in our POMDP formulation. We then compare the proposed POMW policy with its fully observable counterparts. Next, we verify the theoretical analyses on the POMW policy. Finally, we compare the performance of the POMW policy with that of three baseline policies in PONs. All parameter settings are referenced from [26], [27], [28], and [30].

A. Evolution of Belief States

We simulate the evolution of the local ages of a node with the packet arrival rate $\lambda = 0.6$ and $(k,m) \in \{(1,4),(2,3),(3,2),(4,1)\}$ for 50000 runs, and then calculate the distributions of the final values of the local ages on $1,2,3,4,5$ by (21). Let $c_{k,m}(d)$ denote the dth entry in $\boldsymbol{c}(k,m)$. Note that in those cases, the values of the local ages cannot evolve to a value over 5, and hence $c_{k,m}(d) = 0$, $\forall d > 5$. Therefore, we only show the probabilities of values not exceeding 5 for space saving. Table I compares the simulation and theoretical results of these 5 belief states of the local age of a node. The belief states, denoted by $\boldsymbol{c}(k,m)$, are defined in ***Proposition* 1**, i.e., the distribution of the local age of a node with a last observation of the local age k followed by m elapsed slots. According to Table I, the positions of non-zero entries in each belief state of the simulation are the same as that in (21). The simulation results match the theoretical results well, which validates the proposed LOC simplification.

B. Comparisons With Fully Observable Counterparts

The EWSAoI performances of the POMW and FOMW policies are obtained via 2000 Monte-Carlo simulation runs. β_i is set as $\omega_i/\lambda_i\mu_i' p_i$ for both of the two policy. The AoI performance of the RS policies, the universal lower bound L_B, and the upper bound of R^{POMW} are computed using (30), (36), and (33), respectively.

In Fig. 3, we illustrate the EWSAoI of the POMW policy, its corresponding upper bounds, and the universal lower bound L_B with increasing packet arrival rate. The R^{FOMW} in the FON and its corresponding upper bound, i.e., the optimal RS policy, are given as benchmarks. We set $N = 5$, $\omega_i = 1$, $p_i = 0.8$, $\lambda_i = \lambda, \forall i$, and $T = 400$. Fig. 3 shows that all curves decrease as λ increases. This is intuitive because the EWSAoI decreases when the status update packets arrive at nodes more frequently. Furthermore, the value of the universal lower bound is the smallest, the value of R^{RSM} is the largest, and R^{POMW} and R^{FOMW} are lower than their corresponding upper bounds, respectively. These relationships validate the analysis given in the previous section. Fig. 3 also shows that R^{POMW} is larger than R^{FOMW}, and the upper bound of R^{POMW} is larger than that of R^{FOMW}. This is intuitive because the POMW policy in the PON only knows the packet arrival rate and some occasional observations, while the FOMW policy in the FON can utilize the fully observed

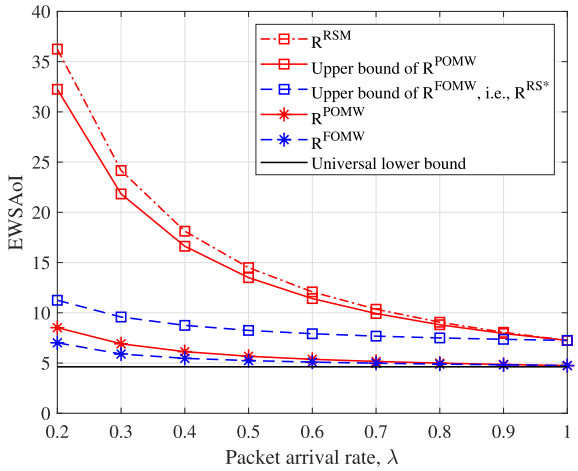


Fig. 3. EWSAoI performance versus an increasing packet arrival rate, where $N = 5$, $p_i = 0.8$, and $\omega_i = 1$.

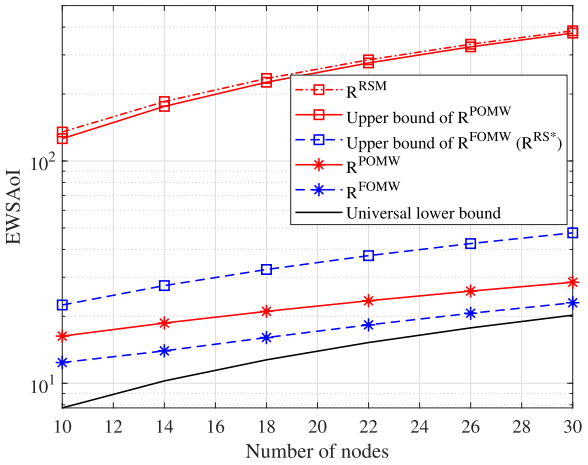


Fig. 5. EWSAoI performance as the number of nodes increases, where $\lambda_i = 0.1$, $p_i = 0.8$, and $\omega_i = 1$.

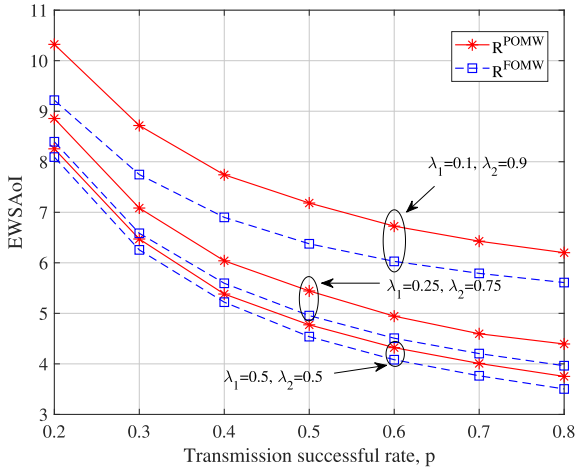


Fig. 4. R^{POMW} and R^{FOMW} versus transmission successful rate with different combinations of λ_i in PON and FON, where $N = 2$, and $\omega_1 = \omega_2 = 1$.

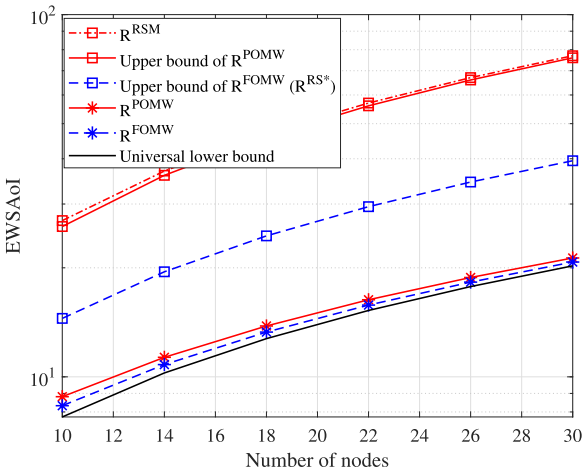
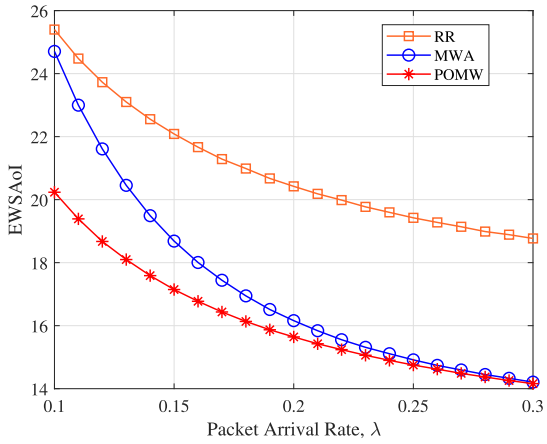


Fig. 6. EWSAoI performance as the number of nodes increases, where $\lambda_i = 0.5$, $p_i = 0.8$, and $\omega_i = 1$.

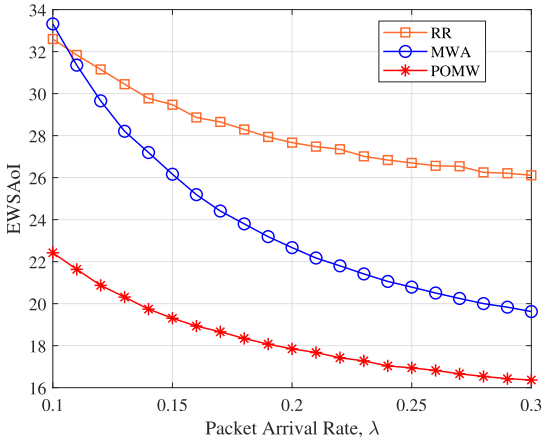
state information. Furthermore, the gap between the R^{FOMW} and its upper bound decreases slowly as λ increases, while the gap between the R^{POMW} and its upper bound decreases much quickly. This phenomenon can be explained by refering to the expressions of the upper bounds, given by (30) and (33). It is obvious that $1/\lambda_i$ in (33) always has larger coefficient than that of (30). Hence, the upper bound of R^{POMW} increases faster than that of R^{FOMW} with the decrease of λ. Moreover, the EWSAoI performances of the POMW and FOMW polices and their corresponding upper bounds converge to the same value when $\lambda = 1$. This is because both the PON and FON approach to the "generate-at-will" model when λ tends to 1.

In Fig. 4, we compare R^{FOMW} and R^{POMW} versus transmission successful rate p for $N = 2$ with three pairs of packet arrival rates $\{\lambda_1, \lambda_2\}$. We set $p_1 = p_2 = p$, $\omega_1 = \omega_2 = 1$, and three pairs of packet arrival rates $\lambda_1 = \lambda_2 = 0.5$; $\lambda_1 = 0.25$, $\lambda_2 = 0.75$; and $\lambda_1 = 0.1$, $\lambda_2 = 0.9$. Fig. 4 shows that the gap between R^{FOMW} and R^{POMW} becomes larger when the gap between two λ_is increase. This can be explained by combining ***Corollary 2*** and [28, Th.5]. ***Corollary 2*** indicates that the performance guarantee of R^{POMW} is inversely proportional to λ_{min}, and [28, Th.5] indicates that the performance guarantee of the fully observed counterpart is a constant. A larger arrival rate gap results in a smaller λ_{min}, and consequently a larger gap between the R^{POMW} and R^{FOMW}. Fig. 4 also shows that R^{POMW} and R^{FOMW} decreases in all cases when the packet transmission rate p increases. This is because the destination AoI $D_{t,i}$ drops to the local age $d_{t,i} + 1$ more frequently with larger p.

In Fig. 5, we depict the EWSAoI of the POMW and FOMW policies with increased number of nodes, and compare them with corresponding upper bounds. We also include R^{RSM} and the universal lower bound as benchmarks. We set $\lambda_i = 0.1$ and $p_i = 0.8$ for all nodes, and increase the number of nodes from 10 to 30. It is shown in Fig. 5 that all curves increase as number of nodes increases. This is because each node has fewer chances to be scheduled when the number of nodes increases, thus its AoI has fewer chances to decrease. Fig. 6 plots the same set of curves as in Fig. 5, where the values of λ_1 and λ_2 are both set to be 0.5. Fig. 6 shows similar phenomenon observed in Fig. 5. Furthermore, the performance



(a) $\omega_i = 1$, $p_i = 0.5$, $\forall i$.



(b) $\omega_i \sim \mathcal{U}(0.1, 1.9)$, $p_i \sim \mathcal{U}(0.1, 0.9)$, $\forall i$.

Fig. 7. EWSAoI performance versus packet arrival rate λ in the PON with $N = 10$. Note that the notation $\mathcal{U}(a, b)$ denotes a uniform distribution over the real number interval $[a, b]$.

of all schemes improves from Fig. 5 to Fig. 6, which is expected since the packet arrival rates are increased.

C. Comparison With Baseline Policies

In the following, we show the advantages of the proposed POMW policy in PONs over two baseline policies. The baseline policies are described as follows:

1) ***Round Robin (RR) policy:*** In the RR policy, nodes are scheduled by the AP in a circular order to ensure a fair scheduling opportunity among the nodes.
2) ***Max weighted AoI (MWA) policy:*** The MWA policy does not need the knowledge of nodes' local age. Specifically, the AP always schedules the node j with $j = \arg\max_i \omega_i p_i D_{t,i}$.

In the following figures, the EWSAoI performance of all policies are obtained via 10000 Monte-Carlo simulation runs.

Fig. 7a shows the EWSAoI performance of the POMW policy, the MWA policy, and the RR policy in a symmetric PON. The weight coefficients ω_i and transmission successful rates p_i of all nodes are the same. We can observe from Fig. 7a that the RR policy has the worst performance. This is intuitive because the RR policy does not use the observations

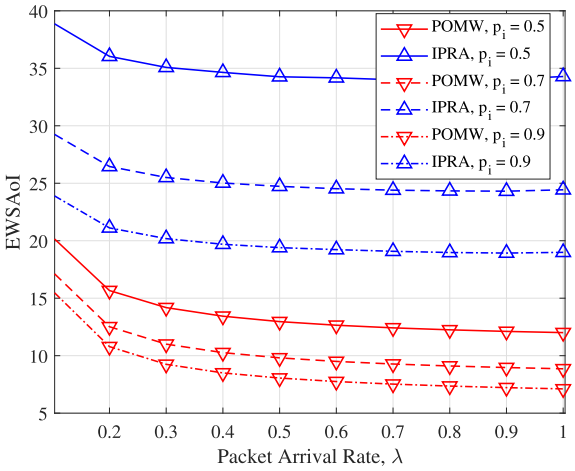


Fig. 8. EWSAoI performance versus packet arrival rate λ, where $N = 10$, and $\omega_i = 1, \forall i$.

of the network states, while the MWA and POMW policies make decisions depending on the observations. Moreover, the POMW policy is superior to the MWA policy when λ is small, but the EWSAoI performances of these two policies tend to coincide when $\lambda \geq 0.25$. This is owing to the fact that the MWA policy only leverages the observations of the destination AoI, while the POMW policy uses the observations of both the destination AoI and local age. Furthermore, in a symmetric PON, the expected local ages of all nodes tend to be symmetric as λ grows. In this case, the POMW policy and the MWA policy becomes equivalent.

Fig. 7b plots the long-term EWSAoI curves of all three policies as in Fig. 7a over asymmetric PONs. The transmission successful rates and weight coefficients of all nodes are randomly drawn from uniform distributions in each simulation run. We can see that in this case, the performance of POMW policy clearly outperforms that of the RR and MWA policies. Furthermore, the performance of the MWA policy cannot approach to that of the POMW policy even when the arrival rate increases to 0.3. This is because the expected local ages are not symmetric in an asymmetric PON and the MWA policy does not consider this information.

Next, we compare the performance of the propose POMW policy with that of a decentralized policy in the considered network. To our best knowledge, the ***Index-Prioritized Random Access (IPRA) policy*** proposed in [27] is the only decentralized scheme that can be applied in the considered system. In the IPRA policy, an index is first calculated for each end node in each slot based on the current local age and AoI of the node, and the end node will transmit only if its index is above a predefined threshold. Fig. 8 shows the EWSAoI performance of the POMW policy and the IPRA policy. We set $N = 10$, $\lambda_i = \lambda$, $\omega_i = 1, \forall i$, and three sets of transmission successful rate with p_i being 0.5, 0.7 and 0.9 for all i, respectively. We can observe from Fig. 8 that the proposed POMW policy is superior to the decentralized IPRA policy in all settings. This is because the decentralized scheme suffers from unavoidable transmission collisions in the uplink.

VII. Conclusion

In this paper, we investigated the AoI-oriented scheduling problem for a wireless multiuser uplink network. Due to the partial observations of the local ages at end devices, we formulated the scheduling decision-making problem as a partially observable Markov decision process (POMDP). The POMDP was first reformulated to an equivalent belief-MDP, and then simplified to an Last-Observation-Characterized (LOC) belief-MDP by adequately leveraging the properties of the status update arrival processes. With the simplification, the infinite dimensional belief states can be characterized by two-dimensional vectors, and thus the complexity of belief updates is significantly reduced. On this basis, we devised the Partially Observable Max-Weight (POMW) policy that minimizes the expected weighted sum AoI of the next slot on condition of the current belief state. Based on the LOC belief-MDP, we derived upper bounds for the performance of the proposed POMW policy. Moreover, we evaluated the performance guarantee of the POMW policy by comparing its performance with a universal lower bound available in the literature. Finally, simulation results validated our analyses, illustrating that the performance gap between the proposed POMW policy and its fully observable counterpart is proportional to the inverse of the lowest arrival rate. The simulation results also validated the superiority of the POMW policy over the baseline policies.

Future work includes the development of a Whittle's index-based policy for the considered scheduling problem, the extension to the scenarios where the packet arrival rates at end nodes are not known a priori, as well as the extension to more recent information freshness metrics (e.g., AoI at Query [44]).

Acknowledgment

The authors would like to thank Tong Zhang and Yijin Zhang for their helpful discussions on establishing the network model and problem formulation.

References

[1] Y. Sun and K. R. Chowdhury, "Enabling emergency communication through a cognitive radio vehicular network," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 68–75, Oct. 2014.

[2] J. Wan et al., "Software-defined industrial Internet of Things in the context of industry 4.0," *IEEE Sensors J.*, vol. 16, no. 20, pp. 7373–7380, Oct. 2016.

[3] R. Talak, S. Karaman, and E. Modiano, "Speed limits in autonomous vehicular networks due to communication constraints," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Dec. 2016, pp. 4998–5003.

[4] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Found. Trends Netw.*, vol. 12, no. 3, pp. 162–259, Nov. 2017.

[5] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the Internet of Things," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 72–77, Dec. 2019.

[6] X. Chen et al., "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2268–2281, Jan. 2020.

[7] J. Liu, X. Wang, B. Bai, and H. Dai, "Age-optimal trajectory planning for UAV-assisted data collection," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2018, pp. 553–558.

[8] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep reinforcement learning for minimizing age-of-information in UAV-assisted networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2019, pp. 1–6.

[9] R. D. Yates, "Age of information in a network of preemptive servers," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2018, pp. 118–123.

[10] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Apr. 2011, pp. 350–358.

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

[12] M. Desai and A. Phadke, "Internet of Things based vehicle monitoring system," in *Proc. 14th Int. Conf. Wireless Opt. Commun. Netw. (WOCN)*, Feb. 2017, pp. 1–3.

[13] Y. Sun, I. Kadota, R. Talak, and E. Modiano, "Age of information: A new metric for information freshness," *Synth. Lectures Commun. Netw.*, vol. 12, no. 2, pp. 1–224, Dec. 2019.

[14] B. Tan Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1122–1126.

[15] Q. Wang, H. Chen, Y. Li, Z. Pang, and B. Vucetic, "Minimizing age of information for real-time monitoring in resource-constrained industrial IoT networks," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2019, pp. 1766–1771.

[16] J. Pan, A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing age of information via scheduling over heterogeneous channels," in *Proc. 22nd Int. Symp. Theory, Algorithmic Found., Protocol Design Mobile Netw. Mobile Comput.*, New York, NY, USA, Jul. 2021, pp. 111–120, doi: 10.1145/3466772.3467040.

[17] 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, Apr. 2016.

[18] Q. He, D. Yuan, and A. Ephremides, "On optimal link scheduling with min-max peak age of information in wireless systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.

[19] B. Dedhia and S. Moharir, "You snooze, you lose: Minimizing channel-aware age of information," in *Proc. 18th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOPT)*, Jun. 2020, pp. 1–8.

[20] 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, Nov. 2017.

[21] S. Leng and A. Yener, "Age of information minimization for an energy harvesting cognitive radio," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 2, pp. 427–439, Jun. 2019.

[22] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 1844–1852.

[23] S. Farazi, A. G. Klein, and D. Richard Brown, "Average age of information for status update systems with an energy harvesting server," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2018, pp. 112–117.

[24] E. Tuğçe Ceran, D. Gündüz, and A. György, "Reinforcement learning to minimize age of information with an energy harvesting sensor with HARQ and sensing cost," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2019, pp. 656–661.

[25] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for age of information for the D/G/1 queue," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2018, pp. 130–135.

[26] 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, Dec. 2018.

[27] J. Sun, Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Closed-form Whittle's index-enabled random access for timely status update," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1538–1551, Mar. 2020.

[28] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Trans. Mobile Comput.*, vol. 20, no. 3, pp. 1173–1185, Mar. 2021.

[29] A. Gong, T. Zhang, H. Chen, and Y. Zhang, "Age-of-information-based scheduling in multiuser uplinks with stochastic arrivals: A POMDP approach," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[30] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Trans. Mobile Comput.*, vol. 19, no. 12, pp. 2903–2915, Dec. 2020.

[31] Z. Chen, N. Pappas, E. Björnson, and E. G. Larsson, "Age of information in a multiple access channel with heterogeneous traffic and an energy harvesting node," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2019, pp. 662–667.

[32] E. Tuğçe Ceran, D. Gündüz, and A. György, "A reinforcement learning approach to age of information in multi-user networks with HARQ," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1412–1426, May 2021.
[33] Q. Wang, H. Chen, C. Zhao, Y. Li, P. Popovski, and B. Vucetic, "Optimizing information freshness via multiuser scheduling with adaptive NOMA/OMA," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1766–1778, Mar. 2022.
[34] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synth. Lectures Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.
[35] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, no. 3, pp. 441–450, Aug. 1987.
[36] G. Yao, A. M. Bedewy, and N. B. Shroff, "Age-optimal low-power status update over time-correlated fading channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 2972–2977.
[37] E. Sert, C. Sönmez, S. Baghaee, and E. Uysal-Biyikoglu, "Optimizing age of information on real-life TCP/IP connections through reinforcement learning," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, May 2018, pp. 1–4.
[38] Y. Shao, Q. Cao, S. C. Liew, and H. Chen, "Partially observable minimum-age scheduling: The greedy policy," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 404–418, Jan. 2022.
[39] D. A. McAllester and S. Singh, "Approximate planning for factored POMDPs using belief state simplification," 2013, *arXiv:1301.6719*.
[40] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. Intell.*, vol. 101, nos. 1–2, pp. 99–134, 1998.
[41] J. Liu, R. Zhang, A. Gong, and H. Chen, "Optimizing age of information in wireless uplink networks with partial observations," 2022, *arXiv:2202.03152*.
[42] S. Wang, L. Huang, and J. Lui, "Restless-UCB, an efficient and low-complexity algorithm for online restless bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 11878–11889.
[43] N. L. Zhang and W. Zhang, "Speeding up the convergence of value iteration in partially observable Markov decision processes," *J. Artif. Intell. Res.*, vol. 14, pp. 29–51, Feb. 2001.
[44] F. Chiariotti et al., "Query age of information: Freshness in pull-based communication," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1606–1622, Mar. 2022.

Jingwei Liu received the B.S. degree in electronic information engineering from Nanjing Tech University, China, the degree in electronic engineering from the Institute of Technology Tallaght, Ireland, in 2015, and the M.S. degree in telecommunications engineering from The University of Sydney, Australia, in 2018. He is currently pursuing the Ph.D. degree with the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include wireless communications and the Internet of Things.

Rui Zhang received the B.S. degree from Southeast University in 2016 and the Ph.D. degree from The University of Sydney in 2020. He is currently a Post-Doctoral Fellow with the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include machine learning in mobile wireless sensing and wireless communications.

Aoyu Gong received the B.S. degree in communication engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2019. He is currently pursuing the M.S. degree with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne. His research interests include design, modeling, and optimization in wireless networks.

He (Henry) Chen (Member, IEEE) received the Ph.D. degree in electrical engineering from The University of Sydney, Sydney, Australia, in 2015. He was a Research Fellow with the School of Electrical and Information Engineering, The University of Sydney. In July 2019, he joined the Department of Information Engineering, The Chinese University of Hong Kong, as a Faculty Member, where he is currently an Assistant Professor. His current research interests include wireless communications, wireless sensing, and their applications in robotic systems. From 2020 to 2022, he served on the editorial board for IEEE WIRELESS COMMUNICATIONS LETTERS. He is serving on the editorial board for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.