AoI-Constrained Transmission Scheduling with HARQ and Heterogeneous Sampling Processes

Saeid Sadeghi Vilni, Mohammad Moltafet, and Markus Leinonen
Centre for Wireless Communications – Radio Technologies
University of Oulu, Finland
e-mail: {saeid.sadeghivilni, mohammad.moltafet, markus.leinonen }@oulu.fi

Marian Codreanu
Department of Science and Technology
Linköping University, Sweden
e-mail: marian.codreanu@liu.se

***Abstract*—We consider a hybrid automatic repeat request (HARQ) based status update system where timely information from multiple sources with different sampling processes (uncontrollable and controllable sampling) is sent via a transmitter to a destination through an error-prone channel. To minimize the average number of transmissions subject to an average age of information constraint, we propose a near-optimal deterministic transmission policy. We formulate a constrained Markov decision process (CMDP) problem and provide a solution using the Lagrangian relaxation method and relative value iteration algorithm. Numerical results show that the proposed policy achieves near-optimal performance.**

Index Terms: AoI, multi-source status update, generate-at-will, random arrival, CMDP.

I. Introduction

There is a growing interest in services that require fresh status update delivery, such as autonomous vehicles, wireless industrial automation, and health monitoring [1]. The Age of Information (AoI) [1]–[3] is a metric used to evaluate the freshness of information in the status update systems. AoI is the difference between the current time and the generation time of the last received packet at a destination [1]–[3].

In the case of an unreliable communication channel, the reliability of data transmission can be enhanced by retransmission protocols [4]. Automatic repeat request (ARQ) protocols are standard error control methods, where after each transmission, the transmitter receives feedback about the reception status of the packet as acknowledgement/negative-acknowledgement (ACK/NACK) [4]. The transmitter keeps retransmitting each packet until it receives an ACK or reaches the maximum allowed number of retransmissions. While ARQ protocols utilize only the last received version of a packet for decoding, hybrid ARQ (HARQ) protocols utilize all received versions to increase the probability of successful decoding the packet [4], [5].

In this paper, we consider a multi-source HARQ-based status update system, where the sources are connected to a transmitter that sends status update packets to a receiver over an unreliable wireless channel (see Fig. 1). We assume a slotted communication, in which the transmitter can send at most one packet per slot. The sources, which monitor some time-varying random processes, are classified into two categories based on their sampling processes: 1) *random arrival* sources (i.e., uncontrollable sampling) which generate status update packets according to a Bernoulli process, and 2) *generate-at-will* sources (i.e., controllable sampling) which can be commanded to generate a status update packets at any slot. Apart from freshness requirements, the radio resources (e.g., power and channel utilization) also play an essential role in the operation of status update systems [6]. Hence, we investigate the problem of minimizing the average number of transmissions subject to the average AoI constraint. The solution of the problem is a policy that determines the transmission status at each time slot, i.e., transmit a fresh packet from a source, retransmit the previously transmitted but not successfully decoded packet from a source, or stay idle. We formulate our problem as a constrained Markov decision process (CMDP) problem. Then, we transform the CMDP problem into an MDP problem via the Lagrangian relaxation method. Using the relative value iteration algorithm (RVIA), we propose a near-optimal practical deterministic transmission policy.

Related Works: AoI characterization has extensively been studied from the perspective of queueing theory; see, e.g., [7]–[9] and the references therein. One of the earliest studies to analyze AoI under an HARQ protocol is [9], where the authors derived the closed-form expression of the average AoI for an HARQ-based M/G/1/1 queueing system. Besides the analysis, the AoI has been studied in the retransmission-based status update systems from the perspective of sampling and transmission policies [10]–[15]. The most related works to our paper are [14], [15]. The work [14] considered a similar HARQ-based status update system to ours, yet with the following differences. The authors in [14] considered a single generate-at-will source, while we consider both random arrival and generate-at-will sources as a multi-source system. Considering the random arrival sources makes the system more complicated, as the transmitter does not know the availability of the fresh packets at the subsequent slots. In [15], which is an extension of [14], the authors considered an HARQ-based status update system that contains one generate-at-will source and several users (destinations), in which at most one user is served at each slot.

This research has been financially supported by the Infotech Oulu, the Academy of Finland (grant 323698), and Academy of Finland 6G Flagship program (grant 346208). The work of M. Leinonen has also been financially supported in part by the Academy of Finland (grant 340171).

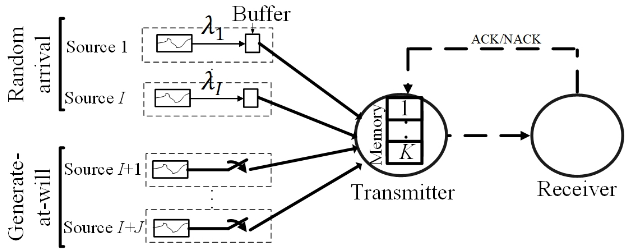

Figure 1. The considered system model.

II. System Model and Problem Formulation

A. System Model

We consider a multi-source status update system that consists of K sources, one transmitter, and one receiver, as depicted in Fig. 1. The receiver is interested in timely information about different random processes monitored by the K sources. The transmitter sends status update packets to the receiver through an error-prone wireless channel with the aid of an HARQ protocol. The system operates in discrete time with unit time slots $t \in \{1, 2, \ldots\}$.

The K sources are divided into two classes based on their sampling processes: 1) a set $\mathcal{I}$ of I *random arrival* sources whose sampling processes are uncontrollable and 2) a set $\mathcal{J}$ of J *generate-at-will* sources, where the transmitter can sample the process at any time. Each source $k \in \mathcal{I}$ generates status update packets randomly and independently at the beginning of slots according to a Bernoulli random process with parameter λ_k. We denote the set of all sources by $\mathcal{K} = \mathcal{I} \cup \mathcal{J} = \{1, \ldots, K\}$, where $K = I + J$.

Each random arrival source has a *buffer* of size one to store the last arrived packet. The transmitter has a *memory* of size K packets to store the previously transmitted but not successfully decoded packets of each source. Note that, after a number of unsuccessful transmission attempts of a packet from a source, the transmitter may decide to transmit a packet from the other sources. In this case, the transmitter retains the previously transmitted but not successfully decoded packet of each source in the transmitter's memory for possible future retransmissions, since this packet is more likely to be decoded than a new packet from that source due to the HARQ protocol. We term a packet in the transmitter's memory an *under-process packet*. Thus, the maximum number of packets stored in the system is $I+K$ packets, i.e., I packets at the buffers of random arrival sources and K packets at the transmitter's memory.

We assume that the transmitter[1] can transmit at most one packet per slot. At each slot, the transmitter decides whether to send a packet or stay idle. The possible transmission options for a random arrival source $k \in \mathcal{I}$ are either transmitting the packet from its buffer or retransmitting the under-process packet from the transmitter's memory. The possible transmission options for a generate-at-will source $k \in \mathcal{J}$ are either generating and transmitting a new sample or retransmitting the under-process packet from the transmitter's memory. We refer to the packets in the buffers of the random arrival sources and to the newly generated packets of the generate-at-will sources as *fresh packets*. If the transmitter decides to transmit a fresh packet from a given source, this packet replaces the source's under-process packet in the transmitter's memory.

1) Transmission Model: At each slot t, the transmitter takes one of the following actions: 1) transmit a fresh packet from a source, 2) retransmit an under-process packet of a source, or 3) stay idle. Let $u_{t,k} \in \{0, 1\}$ denote the decision variable about transmitting a fresh packet from source k at slot t, where $u_{t,k} = 1$ indicates that the transmitter sends the fresh packet, and $u_{t,k} = 0$ otherwise. Let $r_{t,k} \in \{0, 1\}$ denote the decision variable about retransmitting the under-process packet of source k at slot t, where $r_{t,k} = 1$ indicates that the transmitter sends the under-process packet, and $r_{t,k} = 0$ otherwise. Since the transmitter can transmit at most one packet per slot, we have $\sum_{k \in \mathcal{K}} u_{t,k} + r_{t,k} \leq 1$.

HARQ protocol: In the considered HARQ protocol, every packet transmission attempt is followed by an instantaneous error-free ACK/NACK feedback signal from the receiver. Let $d_t \in \{0, 1\}$ denote the packet reception status at slot t, where $d_t = 1$ indicates that the transmitted packet was decoded successfully (ACK), and $d_t = 0$ indicates that either the transmitted packet was not decoded successfully (NACK) or the transmitter remained idle. In the HARQ protocol, the receiver uses all previously received versions of a packet to decode it. Therefore, the probability of successful decoding a packet is an increasing function of the number of attempted transmissions of the packet. Let $x_{t,k}$ denote the number of attempted transmissions of a packet of source k up to slot t. The evolution of $x_{t,k}$ is given as

$$x_{t+1,k} = \begin{cases} 1 & u_{t,k} = 1 \\ x_{t,k} & u_{t,k} + r_{t,k} = 0 \\ x_{t,k} + 1 & r_{t,k} = 1. \end{cases} \tag{1}$$

To account for the fact that most practical HARQ protocols allow only a finite number of retransmissions, we limit the number of transmission attempts of a packet to $x^{\max}$, i.e., $x_{t,k} \leq x^{\max}$. The function representing the probability of successful decoding after $x_{t,k}$ transmissions is denoted by $f(x_{t,k})$. In practice, $f(\cdot)$ is a complicated function of several parameters such as the channel conditions, the channel coding methods, and the combining technique utilized in the HARQ protocol [16], [17].

2) Age of Information: Let $\delta_{t,k}$ denote the AoI of source k at the receiver at slot t; we refer to this simply as the AoI of source k hereinafter. We use the common assumption (see, e.g., [14], [15], [18], [19]) that all AoI values in the system are upper bounded by $\delta^{\max}$. To characterize the AoI of each source, we next define the age of a fresh packet at a source and the age of an under-process packet in the transmitter's memory.

Age of the fresh packets: Let $\delta^{\mathrm{f}}_{t,k}$ denote the age of the fresh packet of source k at slot t. For a random arrival source, if a

[1] A mathematically equivalent system is the one where each source is equipped with an own transmitter while at most one source is allowed to transmit at each slot.

packet arrives at the buffer at the beginning of slot t, the age of the fresh packet becomes zero, otherwise it is incremented by one. Let $b_{t,k} \in \{0,1\}$ denote the packet arrival status of source $k \in \mathcal{I}$ at slot t, where $b_{t,k} = 1$ indicates a packet arrives at the buffer, and $b_{t,k} = 0$ otherwise. Note that $\Pr(b_{t,k} = 1) = \lambda_k$. For the generate-at-will sources, the transmitter can generate a fresh packet at any time so that the age of the fresh packet is always zero. Thus, the evolution of $\delta^{\mathrm{f}}_{t,k}$ with the initial value $\delta^{\mathrm{f}}_{0,k} = 0$ is given as

$$\delta^{\mathrm{f}}_{t,k} = \begin{cases} 0 & b_{t,k} = 1,\ k \in \mathcal{I} \\ \min\{\delta^{\mathrm{f}}_{t-1,k} + 1, \delta^{\max}\} & b_{t,k} = 0,\ k \in \mathcal{I} \\ 0 & k \in \mathcal{J}, \end{cases} \tag{2}$$

Age of the under-process packets: Let $\delta^{\mathrm{p}}_{t,k}$ denote the age of the under-process packet of source k at slot t. If the transmitter sends a fresh packet of source k at slot t, the age of the under-process packet of the source at the next slot drops to $\min\{\delta^{\mathrm{f}}_{t,k}+1, \delta^{\max}\}$. In other cases (i.e., retransmission or staying idle), the age of the under-process packet is incremented by one. The evolution of $\delta^{\mathrm{p}}_{t,k}$ with the initial value $\delta^{\mathrm{p}}_{0,k} = 0$ is given by

$$\delta^{\mathrm{p}}_{t+1,k} = \begin{cases} \min\{\delta^{\mathrm{f}}_{t,k} + 1, \delta^{\max}\} & u_{t,k} = 1 \\ \min\{\delta^{\mathrm{p}}_{t,k} + 1, \delta^{\max}\} & \text{otherwise.} \end{cases} \tag{3}$$

AoI at the receiver: We now characterize the evolution of the AoI at the receiver. If the transmitter sends a fresh packet of source k at slot t (i.e., $u_{t,k} = 1$) and the packet is decoded successfully at the receiver (i.e., $d_t = 1$), the AoI of the source at the next slot drops to $\min\{\delta^{\mathrm{f}}_{t,k} + 1, \delta^{\max}\}$, otherwise (i.e., $d_t = 0$), the AoI increases by one. If the transmitter retransmits the under-process packet of source k (i.e., $r_{t,k} = 1$) and it is decoded successfully at the receiver, the AoI of the source at the next slot drops to $\min\{\delta^{\mathrm{p}}_{t,k} + 1, \delta^{\max}\}$, otherwise (i.e., $d_t = 0$), the AoI increases by one. If, at slot t, the transmitter does not transmit any packet of source k (i.e., $u_{t,k}+r_{t,k} = 0$), the AoI of the source at the next slot increases by one. The evolution of $\delta_{t,k}$ with the initial value $\delta_{0,k} = 0$ is given as

$$\delta_{t+1,k} = \begin{cases} \min\{\delta^{\mathrm{f}}_{t,k} + 1, \delta^{\max}\} & u_{t,k} d_t = 1 \\ \min\{\delta^{\mathrm{p}}_{t,k} + 1, \delta^{\max}\} & r_{t,k} d_t = 1 \\ \min\{\delta_{t,k} + 1, \delta^{\max}\} & u_{t,k}(1 - d_t) = 1 \\ \min\{\delta_{t,k} + 1, \delta^{\max}\} & r_{t,k}(1 - d_t) = 1 \\ \min\{\delta_{t,k} + 1, \delta^{\max}\} & u_{t,k} + r_{t,k} = 0. \end{cases} \tag{4}$$

B. Problem Formulation

Our main goal is to minimize the average number of transmissions subject to the average AoI constraint by finding a transmission policy that determines the transmission decision variables at each slot t.

Let $\tau_t \in \{0,1\}$ denote the transmission status at slot t, where $\tau_t = 1$ indicates that the transmitter sends a packet, and $\tau_t = 0$ otherwise. Thus, we have

$$\tau_t = \begin{cases} 1 & \sum_{k\in\mathcal{K}} u_{t,k} + r_{t,k} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Let $\bar{\tau}$ denote the expected long-term time average number of transmissions, defined as

$$\bar{\tau} = \limsup_{T\to\infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \mathbb{E}\{\tau_t\}, \tag{6}$$

where $\mathbb{E}\{\cdot\}$ is the expectation with respect to the randomness of the system (i.e., packet arrival processes of the random arrival sources and randomness in the communication channel) and the decision variables $\{u_{t,k}, r_{t,k}\}_{k\in\mathcal{K}}$. Finally, let $\bar{\delta}$ denote the expected long-term time average of AoI, given as

$$\bar{\delta} = \limsup_{T\to\infty} \tfrac{1}{TK} \textstyle\sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{E}\{\delta_{t,k}\}, \tag{7}$$

The problem is formulated as a constrained Markov decision process (CMDP) problem. The CMDP is defined by a tuple of five elements $(\mathcal{S}, \mathcal{A}_s, \mathcal{P}, c, d)$: state space, action space, state transition probabilities, and two cost functions, which are defined in the following.

State: Let $s_{t,k} = \{\delta^{\mathrm{f}}_{t,k}, \delta^{\mathrm{p}}_{t,k}, \delta_{t,k}, x_{t,k}\}$ denote the state of source k at slot t. The system state at slot t is defined as $s_t = \{s_{t,k}\}_{k\in\mathcal{K}} \in \mathcal{S}$, where $\mathcal{S}$ is the state space. The initial state is denoted with $s_0 = \{s_{0,k}\}_{k\in\mathcal{K}}$, where $s_{0,k} = \{0,0,0,0\}$ for all $k \in \mathcal{K}$.

Action: Let $a_t = \{a_{t,k}\}_{k\in\mathcal{K}} \in \mathcal{A}_{s_t}$ denote the action of the transmitter at slot t, where $a_{t,k} = \{u_{t,k}, r_{t,k}\}$ represents the action for source k, and $\mathcal{A}_{s_t}$ is a space of feasible actions in state s_t, defined as $\mathcal{A}_{s_t} = \big\{u_{t,k}, r_{t,k} \in \{0,1\} \mid k \in \mathcal{K},\ \sum_{k\in\mathcal{K}} u_{t,k} + r_{t,k} \le 1,\ r_{t,k}(x_{t,k}+1) \le x^{\max}\big\}$.

Cost functions: The CMDP has two cost functions: 1) transmission cost, defined as $c(a_t) = \tau_t$, i.e., $c(a_t) = 1$ if the transmitter makes a transmission attempt at slot t, otherwise $c(a_t) = 0$, and 2) AoI cost, defined as $e(s_t) = \frac{1}{K}\sum_{k=1}^{K} \delta_{t,k}$.

State transition probabilities: Let $\mathcal{P}(s' \mid s, a)$ denote the state transition probabilities, defined as the probability of moving from current state s to a next state s' under action a. Given an action, the one-slot evolution of the AoI values (at the source, memory, and destination) and of the number of transmissions of the under-process packets is independent among the sources. Therefore, the state transition probability factorizes as $\mathcal{P}(s' \mid s, a) = \Pi_{k\in\mathcal{K}} \Pr(s'_k \mid s_k, a_k)$. Let us denote $\tilde{\delta}^{\mathrm{f}}_k \triangleq \min\{\delta^{\mathrm{f}}_k + 1, \delta^{\max}\}$, $\tilde{\delta}^{\mathrm{p}}_k \triangleq \min\{\delta^{\mathrm{p}}_k + 1, \delta^{\max}\}$, $\tilde{\delta}_k \triangleq \min\{\delta_k + 1, \delta^{\max}\}$, $\bar{f}(\cdot) \triangleq 1 - f(\cdot)$, and $\bar{\lambda}_k \triangleq 1 - \lambda_k$. Given the state $s_k = \{\delta^{\mathrm{f}}_k, \delta^{\mathrm{p}}_k, \delta_k, x_k\}$, the state transition probabilities for a random arrival source $k \in \mathcal{I}$ for the non-zero cases are given by

$$\Pr\big(\{0, \tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}^{\mathrm{f}}_k, 1\} \mid s_k, a_k = \{1,0\}\big) = f(1)\lambda_k \tag{8a}$$

$$\Pr\big(\{\tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}^{\mathrm{f}}_k, 1\} \mid s_k, a_k = \{1,0\}\big) = f(1)\bar{\lambda}_k \tag{8b}$$

$$\Pr\big(\{0, \tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}_k, 1\} \mid s_k, a_k = \{1,0\}\big) = \bar{f}(1)\lambda_k \tag{8c}$$

$$\Pr\big(\{\tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}_k, 1\} \mid s_k, a_k = \{1,0\}\big) = \bar{f}(1)\bar{\lambda}_k \tag{8d}$$

$$\begin{aligned}\Pr\big(\{0, \tilde{\delta}^{\mathrm{p}}_k, \tilde{\delta}^{\mathrm{p}}_k, x_k+1\} \mid s_k, a_k = \{0,1\}\big)& \\ = f(x_k+1)\lambda_k&\end{aligned} \tag{8e}$$

$$\begin{aligned}\Pr\big(\{\tilde{\delta}^{\mathrm{f}}_k, \tilde{\delta}^{\mathrm{p}}_k, \tilde{\delta}^{\mathrm{p}}_k, x_k+1\} \mid s_k, a_k = \{0,1\}\big)& \\ = f(x_k+1)\bar{\lambda}_k&\end{aligned} \tag{8f}$$

$$\Pr\big(\{0,\tilde{\delta}_k^{\mathrm{p}},\tilde{\delta}_k,x_k+1\} \mid s_k,a_k=\{0,1\}\big) \\ =\bar{f}(x_k+1)\lambda_k \tag{8g}$$

$$\Pr\big(\{\tilde{\delta}_k^{\mathrm{f}},\tilde{\delta}_k^{\mathrm{p}},\tilde{\delta}_k,x_k+1\} \mid s_k,a_k=\{0,1\}\big) \\ =\bar{f}(x_k+1)\bar{\lambda}_k \tag{8h}$$

$$\Pr\big(\{0,\tilde{\delta}_k^{\mathrm{p}},\tilde{\delta}_k,x_k\} \mid s_k,a_k=\{0,0\}\big)=\lambda_k \tag{8i}$$

$$\Pr\big(\{\tilde{\delta}_k^{\mathrm{f}},\tilde{\delta}_k^{\mathrm{p}},\tilde{\delta}_k,x_k\} \mid s_k,a_k=\{0,0\}\big)=\bar{\lambda}_k, \tag{8j}$$

The state transition probabilities for the generate-at-will source $k\in\mathcal{J}$ are obtained by substituting $\lambda_k=1$ in (8).

Let π denote a policy that determines the action taken at each state. A stationary randomized policy is mapping from each state to a distribution over actions, $\pi(a\mid s):\mathcal{S}\times\mathcal{A}\rightarrow[0,1],\ \sum_{a\in\mathcal{A}_s}\pi(a\mid s)=1$. A (stationary) deterministic policy chooses an action at a given state with probability one, which is a special case of the stationary randomized policy. With a slight abuse of notation, we denote the action taken in state s by a deterministic policy π with $\pi(s)$. Let $\bar{\tau}^{\pi}=\limsup\limits_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{c(a_t)\mid s_0\}$ denote the average number of transmissions (see (6)), obtained under policy π starting from the initial state s_0. Let $\bar{\delta}^{\pi}=\limsup\limits_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{e(s_t)\mid s_0\}$ denote average AoI (see (7)), obtained under policy π starting from the initial state s_0. Having constructed the CMDP, the CMDP problem is given as

$$\begin{array}{ll}\underset{\pi}{\text{minimize}} & \bar{\tau}^{\pi}\\ \text{subject to} & \bar{\delta}^{\pi}\leq\Delta^{\max},\end{array} \tag{9}$$

where $\Delta^{\max}$ is the maximum allowed average AoI. An optimal policy that solves CMDP problem (9) is denoted with π^*, and the optimal value of the problem is denoted with $\bar{\tau}^*$.

Similarly to [20], [21], to solve problem (9), we need to make extra assumptions about the CMDP structure. Specifically, we assume that given the initial state (s_0), all policies will induce a Markov chain with only one recurrent class and a (possibly empty) set of transient states. This assumption makes problem (9) well-posed so that we can use the tools associated with the unichain MDPs, as described in the next section.

III. Deterministic Transmission policy

In this section, we propose a (near-optimal) solution to the CMDP problem. To this end, we apply the Lagrangian relaxation method to transform the CMDP problem to an (unconstrained) MDP problem, parametrized by a Lagrange dual variable [22, Sec. 3.3]. In comparison to the CMDP problem, the MDP problem has only one cost function that is defined as $L(s,a,\beta)=c(a_t)+\beta e(s_t)$, whereas the other elements, i.e., the state space, action space, and state transition probabilities, are the same. Let $\bar{L}(\pi,\beta)=\limsup_{T\rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{c(a_t)+\beta\big(e(s_t)-\Delta^{\max}\big)\}$ denote the Lagrangian corresponding to CMDP problem (9), where β is the Lagrangian multiplier. Following the standard Lagrangian relaxation procedure, we restrict to the set of deterministic policies and construct the following MDP problem associated with the CMDP problem (9)

$$\underset{\pi\in\Pi_{\mathrm{D}}}{\text{minimize}}\quad \bar{L}(\pi,\beta), \tag{10}$$

where Π_{D} is the set of all deterministic policies. Let π_{β}^* denote an optimal policy that solves problem (10) for a given β, which is called a β-*optimal* policy.

The following remark expresses the relation between the optimal values of CMDP problem (9) and the MDP problem (10).

Remark 1: The cost function in the objective of CMDP problem (9) is bounded below, i.e., $c(a_t)\geq 0$ for all $t\in\mathbb{N}$. Moreover, the state space, $\mathcal{S}$, is finite. Therefore, the two conditions in [22, Corollary 12.2] are satisfied in our CMDP formulation, and we have

$$\bar{\tau}^*=\sup_{\beta\geq 0}\min_{\pi\in\Pi_{\mathrm{D}}}\bar{L}(\pi,\beta). \tag{11}$$

According to Remark 1, the optimal value of CMDP problem (9), $\bar{\tau}^*$, is obtained via the solution of the right hand side of (11), which means finding the optimal Lagrangian multiplier β^* and its corresponding β^*-optimal policy, $\pi_{\beta^*}^*$. If policy $\pi_{\beta^*}^*$ satisfies the constraint of CMDP problem (9) with equality, i.e., $\bar{\delta}^{\pi_{\beta^*}^*}=\Delta^{\max}$, then $\pi_{\beta^*}^*$ is an optimal policy for the CMDP problem, i.e., $\pi^*=\pi_{\beta^*}^*$. However, due to the discrete nature of the action space, in general, there is no guarantee that $\pi_{\beta^*}^*$ satisfies the constraint with equality. To elaborate this further, the following remark presents the structure of an optimal policy π^*.

Remark 2: An optimal policy for CMDP problem (9), π^*, is a randomized mixture of two deterministic $\tilde{\beta}$-optimal policies, from which one policy satisfies the constraint and the other one violates it. The two policies are mixed with a randomization factor such that the obtained optimal policy satisfies $\bar{\delta}^{\pi^*}=\Delta^{\max}$ [14], [23].

Finding the optimal policy becomes readily computationally intractable even for a moderate number of states, especially because oftentimes, the optimal randomization factor can be found only numerically [24, Section 3.2]. Therefore, in order to solve CMDP problem (9), we propose a practical deterministic policy, which is numerically shown to provide near-optimal performance in Section IV. More specifically, we develop an iterative algorithm based on bisection and the relative value iteration algorithm (RVIA), as summarized in Algorithm 1. In brief, at each iteration, we find a β-optimal policy for a given β via the RVIA and subsequently update β according to the bisection rule. The iterative procedure continues until the best β-optimal policy among the feasible β-optimal policies is found. In the next two subsections, we delve into details of this procedure.

1) Algorithm to Find a β-optimal Policy: To obtain an optimal policy π_{β}^* for a given β, we solve the MDP problem (10) via RVIA. By [25, Theorem 8.4.3], there exists a relative value function $h(s)$, for all $s\in\mathcal{S}$, that satisfies

$$\bar{L}^*(\beta)+h(s) \\ =\min_{a\in\mathcal{A}_s}\big[L(s,a,\beta)+\textstyle\sum_{s'\in\mathcal{S}}\Pr(s'\mid s,a)h(s')\big],$$

where $\bar{L}^*(\beta)$ is the optimal value of the MDP problem (10) for a given β, defined as $\bar{L}^*(\beta)=\min_{\pi\in\Pi_{\mathrm{D}}}\bar{L}(\pi,\beta)$.

Subsequently, the β-optimal policy, π^*_β, is obtained as [25, Theorem 8.4.4]

$$\pi^*_\beta(s) = \arg\min_{a\in\mathcal{A}_s}\left[L(s,a,\beta)+\sum\nolimits_{s'\in\mathcal{S}}\Pr(s'\mid s,a)h(s')\right]. \tag{12}$$

To obtain the β-optimal policy, we use the RVIA, in which the relative value function for all states $s\in\mathcal{S}$ at each iteration $i\in\{0,1,\dots\}$ is updated as $h^i(s) = v^i(s) - v^i(s^{\text{ref}})$, where $s^{\text{ref}}\in\mathcal{S}$ is an arbitrary reference state which remains unchanged throughout the iterations. The term $v^i(s)$, called value function, is obtained at each iteration as $v^i(s) = \min_{a\in\mathcal{A}_s}\left[L(s,a,\beta)+\sum_{s'\in\mathcal{S}}\Pr(s'\mid s,a)h^{i-1}(s')\right]$.

For any state $s\in\mathcal{S}$ and initialization $v^0(s)$, the sequences $\{h^i(s)\}_{i=1,2,\dots}$ and $\{v^i(s)\}_{i=1,2,\dots}$ converge, i.e., $\lim_{i\to\infty}h^i(s) = h(s)$ and $\lim_{i\to\infty}v^i(s) = v(s)$. The RVI algorithm to find a β-optimal policy is presented in Steps 3-10 of Algorithm 1. After the convergence of RVIA, i.e., the convergence of the relative value function, $h(\cdot)$, and the value function, $v(\cdot)$ (see Steps 4-8 in Algorithm 1), we obtain the β-optimal policy, π^*_β, according to (12) (see Steps 9-10 in Algorithm 1). It is worth noting that the optimal value of the MDP problem (10) for a given β is given by $\bar{L}^*(\beta) = v(s^{\text{ref}})$.

2) Algorithm to Find the Optimal Lagrangian Multiplier: According to [23, Lemma 3.1], for a given β-optimal policy (π^*_β), the objective function of the CMDP problem, $\bar{\tau}^{\pi^*_\beta}$, and the objective function of the MDP problem, $\bar{L}^*(\beta)$, are increasing in β, while the constraint of the CMDP problem, $\bar{\delta}^{\pi^*_\beta}$, is decreasing in β. Therefore, we are interested in the smallest Lagrangian multiplier that satisfies the constraint in CMDP problem (9), defined as $\tilde{\beta}\triangleq\inf\{\beta\geq 0\mid\bar{\delta}^{\pi^*_\beta}\leq\Delta^{\max}\}$.

To search for $\tilde{\beta}$, we use the bisection algorithm which takes advantage of the monotonicity of $\bar{\delta}^{\pi^*_\beta}$ with respect to β, as presented in Algorithm 1 (see Steps 1-14). We initialize the bisection algorithm with β_{u} and β_{l} in such a way that $\bar{\delta}^{\pi^*_{\beta_{\text{u}}}}\leq\Delta^{\max}$ and $\bar{\delta}^{\pi^*_{\beta_{\text{l}}}}\geq\Delta^{\max}$, which also implies $\beta_{\text{u}}\geq\beta_{\text{l}}$. The algorithm termination criterion is $\beta_{\text{u}}-\beta_{\text{l}}<\kappa$, where κ is a sufficiently small constant. After termination of the bisection algorithm, we set $\tilde{\beta}=\beta_{\text{u}}$ and obtain the best feasible β-optimal policy as $\pi^*_{\tilde{\beta}} = \pi^*_{\beta_{\text{u}}}$. Moreover, the algorithm returns the infeasible policy associated with β_{l}, which represents a lower-bound to an optimal solution of (9).

Algorithm 1: The deterministic transmission policy

Input: $\Delta^{\max}$, $f(\cdot)$, $\lambda_k\ \forall\ k\in\mathcal{I}$, s^{ref}, ϵ, β_{u}, β_{l}, and κ

1 **while** $\beta_{\text{u}}-\beta_{\text{l}}\geq\kappa$ **do**
2 $\quad\bar{\beta}=\frac{\beta_{\text{u}}+\beta_{\text{l}}}{2}$
3 $\quad$**Initialize:** $i=1$, $h^0(s)=1$, $h^1(s)=0$, $v^0(s)=0$ $\forall\ s\in\mathcal{S}$
4 $\quad$**while** $\max_{s\in\mathcal{S}}|h^i(s)-h^{i-1}(s)|\geq\epsilon$ **do**
5 $\qquad i=i+1$
6 $\qquad$**for** $s\in\mathcal{S}$ **do**
7 $\qquad\quad v^i(s)=\min_{a\in\mathcal{A}_s}\left[L(s,a,\bar{\beta})+\sum_{s'\in\mathcal{S}}\Pr(s'\mid s,a)h^{i-1}(s')\right]$
8 $\qquad\quad h^i(s)=v^i(s)-v^i(s^{\text{ref}})$
9 $\quad$**for** $s\in\mathcal{S}$ **do**
10 $\qquad\pi^*_{\bar{\beta}}(s)=\arg\min_{a\in\mathcal{A}_s}\left[L(s,a,\bar{\beta})+\sum_{s'\in\mathcal{S}}\Pr(s'\mid s,a)h^i(s')\right]$
11 $\quad$**if** $\bar{\delta}^{\pi^*_{\bar{\beta}}}\leq\Delta^{\max}$ **then**
12 $\qquad\beta_{\text{u}}=\bar{\beta}$
13 $\quad$**else**
14 $\qquad\beta_{\text{l}}=\bar{\beta}$

Output: Feasible policy: $\pi^*_{\beta_{\text{u}}}$, infeasible policy: $\pi^*_{\beta_{\text{l}}}$

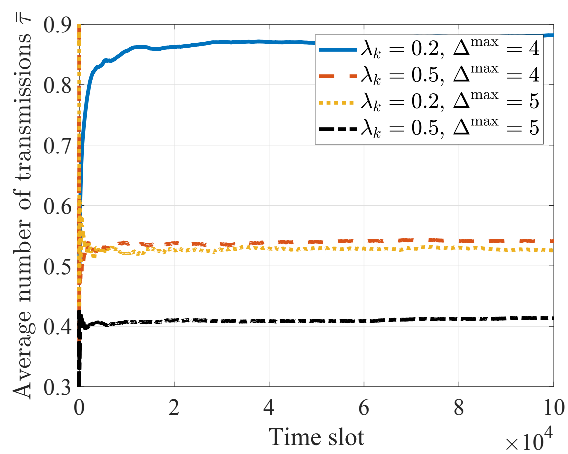

Figure 2. The evolution of the average number of transmissions, $\bar{\tau}$, versus time slots for different packet arrival rate λ_k for $k\in\mathcal{I}$.

IV. Numerical Results

In this section, we evaluate the performance of the proposed transmission scheduling policy. For the probability of successful decoding, we use the function in [14], i.e., $f(x_{t,k}) = 1 - p_0\eta^{x_{t,k}-1}$, where $p_0\in[0,1]$ is the error probability of the first transmission of a packet and $\eta\in[0,1]$ determines the effectiveness of the HARQ protocol. We consider one random arrival source and one generate-at-will source, i.e., $K=2$, and we set $\eta=0.4$, $p_0=0.4$, $\delta^{\max}=18$, and $x^{\max}=5$. The rest of the parameters are specified in each figure. We set the bounds on the Lagrangian multiplier as $\beta_{\text{u}}=1$, $\beta_{\text{l}}=0$, the bisection stopping criterion as $\kappa=0.005$, and the RVIA stopping criterion as $\epsilon=0.01$.

Fig. 2 shows the evolution of $\bar{\tau}$ with respect to time slots for the different packet arrival rate, λ_k, and $\Delta^{\max}$. From Fig. 2, it can be seen that when λ_k decreases, the average number of transmissions increases dramatically. For example, when $\Delta^{\max}=4$, by decreasing λ_k from 0.5 to 0.2, the value of $\bar{\tau}$ increases by about 75 %. This is because when λ_k decreases, the availability of the fresh packets at the random arrival source decreases, and consequently, the AoI of this source increases. In this case, to satisfy the AoI constraint, the transmitter must send the generate-at-will source's packets more frequently to compensate for the negative effect of the random arrival sources on the average AoI.

Fig. 3 shows the average number of transmissions, $\bar{\tau}$, as a function of $\Delta^{\max}$ under (feasible) proposed policy and the (infeasible) lower-bound policy, obtained in Algorithm 1. Furthermore, we consider a (feasible) baseline policy, where

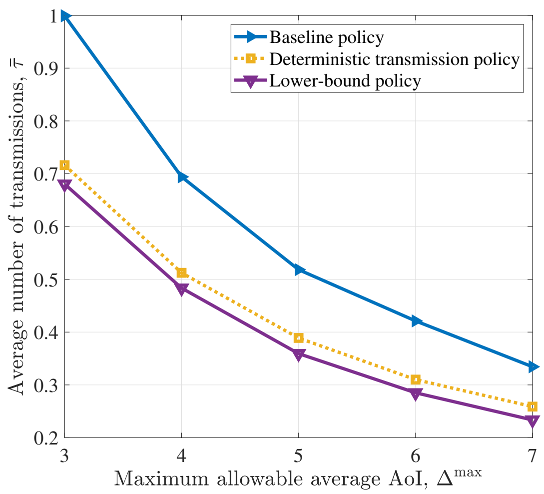

Figure 3. The average number of transmissions, $\bar{\tau}$, for the proposed transmission policy versus $\Delta^{\max}$ where $\lambda_k = 0.7$ for all $k \in \mathcal{I}$.

the transmitter sends a packet whenever the average AoI reaches $\Delta^{\max}$. In every transmission attempt, the source with larger AoI is selected; if there are multiple sources with the largest AoI, one of them is selected randomly. The policy employs an HARQ protocol where the transmitter persistently re-transmits the packet at consecutive slots until it is transmitted successfully or reaches the maximum allowed number of transmissions $x^{\max}$. According to Fig. 3, as expected, the lower-bound policy outperforms the proposed policy as it does not satisfy the constraint. The deterministic transmission policy has a small gap with the lower-bound policy, which implies its near-optimal performance. In general, compared to the baseline policy, the proposed policy improves the system performance considerably, e.g., the policy provides about $40\ \%$ improvement.

V. Conclusion

We studied an HARQ-based multi-source status update system with random arrival and generate-at-will sources, communicating through an error-prone channel. We solved the problem of minimizing the average number of transmissions subject to the average AoI constraint. We developed a deterministic transmission policy using the RVIA and the bisection. Numerical results showed the near-optimal performance of the deterministic transmission policy. Overall, the results showed about $40\ \%$ performance gain for the proposed policy over a baseline scheduling policy.

References

[1] 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.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE Int. Conf. on Computer Commun. (INFOCOM)*, Orlando, FL, USA, Mar. 25–30, 2012, pp. 2731–2735.

[3] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. Conf. Inform. Sciences Syst. (CISS)*, Princeton, NJ, USA, Mar.21–23, 2012, pp. 1–6.

[4] S. Lin, D. J. Costello, and M. J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, vol. 22, no. 12, pp. 5–17, Dec. 1984.

[5] "IEEE standard for air interface for broadband wireless access systems," *IEEE Std 802.16-2017 (Revision of IEEE Std 802.16-2012)*, pp. 1–2726, Mar. 2018.

[6] V. Raghunathan, C. Schurgers, S. Park, and M. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 40–50, Mar. 2002.

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

[8] M. Moltafet, M. Leinonen, and M. Codreanu, "Average AoI in multi-source systems with source-aware packet management," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1121–1133, Feb. 2021.

[9] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *Proc. IEEE Int. Symp. Inform. Theory*, Aachen, Germany, Jun. 25–30, 2017, pp. 131–135.

[10] Y. Shi, L. Jing, X. Jia, P. Ji, and N. Wan, "Improvement on age of information for information update systems with HARQ chase combining and sensor harvesting-transmitting diversities," *IEEE Access*, vol. 9, pp. 78 035–78 049, May 2021.

[11] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Timely status updating over erasure channels using an energy harvesting sensor: Single and multiple sources," *IEEE Trans. Green Commun. Net.*, vol. 6, no. 1, pp. 6–19, Mar. 2022.

[12] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: Without and with feedback," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5091–5105, May 2021.

[13] Y. Wang, S. Wu, J. Jiao, W. Wu, Y. Wang, and Q. Zhang, "Age-optimal transmission policy with HARQ for freshness-critical vehicular status updates in space-air-ground integrated networks," *IEEE Internet of Things J.*, vol. 9, no. 8, pp. 5719–5729, Apr. 2022.

[14] E. T. Ceran, D. Gündüz, and A. György, "Average age of information with hybrid ARQ under a resource constraint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1900–1913, Mar. 2019.

[15] E. T. Ceran, D. Gündüz, and A. György, "A reinforcement learning approach to age of information in multi-user networks with HARQ," *IEEE J. Select. Areas Commun.*, vol. 39, no. 5, pp. 1412–1426, May 2021.

[16] X. Lagrange, "Throughput of HARQ protocols on a block fading channel," *IEEE Commun. Lett.*, vol. 14, no. 3, pp. 257–259, Mar. 2010.

[17] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," in *Proc. IEEE Veh. Technol. Conf.*, Atlantic City, NJ, USA, Oct. 2001, pp. 1829–1833.

[18] M. Moltafet, M. Leinonen, M. Codreanu, and N. Pappas, "Power minimization for age of information constrained dynamic control in wireless sensor networks," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 419–432, Jan. 2022.

[19] A. Maatouk, S. Kriouile, M. Assad, and A. Ephremides, "On the optimality of the Whittle's index policy for minimizing the age of information," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1263–1277, Oct. 2021.

[20] G. Yao, A. Bedewy, and N. B. Shroff, "Age-optimal low-power status update over time-correlated fading channel," *IEEE Trans. Mobile Comput.*, pp. 1–1, Mar. 2022.

[21] D. V. Djonin and V. Krishnamurthy, "MIMO transmission control in fading channels—a constrained Markov decision process formulation with monotone randomized policies," *IEEE Transactions on Signal processing*, vol. 55, no. 10, pp. 5069–5083, 2007.

[22] E. Altman, *Constrained Markov decision processes*. CRC Press, 1999.

[23] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint," *Journal of mathematical analysis and applications*, vol. 112, no. 1, pp. 236–252, Nov. 1985.

[24] D.-j. Ma, A. M. Makowski, and A. Shwartz, "Estimation and optimal control for constrained Markov chains," in *1986 25th IEEE Conf. on Decision and Contr.*, Athens, Greece, Dec. 1986, pp. 994–999.

[25] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.