

# Minimizing AoI in Resource-Constrained Multi-Source Relaying Systems with Stochastic Arrivals

Abolfazl Zakeri, Mohammad Moltafet, and Markus Leinonen

Centre for Wireless Communications – Radio Technologies

University of Oulu, Finland

e-mail: {abolfazl.zakeri, 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 multi-source relaying system where the sources independently and randomly generate status update packets which are sent to the destination with the aid of a buffer-aided relay through unreliable links. We formulate a stochastic optimization problem aiming to minimize the sum average age of information (AAoI) of sources under per-slot transmission capacity constraints and a long-run average resource constraint. To solve the problem, we recast it as a constrained Markov decision process (CMDP) problem and adopt the Lagrangian method. We analyze the structure of an optimal policy for the resulting MDP problem that possesses a switching-type structure. We propose an algorithm that obtains a stationary deterministic near-optimal policy, establishing a benchmark for the system. Simulation results show the effectiveness of our algorithm compared to benchmark algorithms.

**Index Terms**—Age of information (AoI), multi-source scheduling, relay, constrained Markov decision process (CMDP), Lagrangian method.

## I. INTRODUCTION

In many emerging applications such as the Internet-of-Things, cyber-physical systems, and intelligent transportation systems, the freshness of status information is crucial [1]. The *age of information* (AoI) has been proposed to characterize the information freshness in status update systems [2], [3]. The AoI is defined as the time elapsed since the latest received status update packet was generated. Recently, the AoI has attracted much interest in different areas, e.g., queuing systems [4], [5] and scheduling problems [6]–[9].

In some status update systems, there is no direct communication link between the source of information and the intended destination or direct communication is costly. Deploying an intermediate node, typically called a *relay*, in such systems is indispensable and has an array of benefits, e.g., saving on power usage of wireless sensors and improving the transmission success probability. Recently, the AoI has been studied in relaying systems in [6], [10]–[14]. The work [6] studied the AoI minimization in a multi-source relaying system with the *generate-at-will* model (i.e., possibility of generating a new update in any time), under a transmission capacity constraint for each link. They provided an optimal scheduling policy for a setting called the error-prone symmetric IoT network. The authors in [10] analyzed the AoI in a discrete-time Markovian system for two different relay settings and analyzed the impact of the relay on the AoI. In [11], the authors analyzed the

average AoI (AAoI) in a two-way relaying system in which two sources exchange status data, considering the generate-at-will model. The AoI performance under different policies (e.g., a last-generated-first-served policy) in a general multi-hop networks with single-source was studied in [12]. In [13], the authors studied the AoI in a single-source energy harvesting relaying system with error-free channels, and they designed offline and online age-optimal policies. In [14], the authors considered a single-source relaying system with stochastic arrivals where the source communicates with the destination either through the direct link or via a relay. They proposed two different relaying protocols and analyzed the AoI performance.

At the same time, resource limitations are a main bottleneck in optimizing AoI in status update systems, especially, in power-limited sensor networks. Only a few works, such as [13] has incorporated a resource constraint in the relaying system in analysis of the AoI. Moreover, most of the discussed works consider single-source relaying systems, e.g., [10], [12]–[14]. Accordingly, the AoI in *multi-source resource-constrained* relaying systems has not been widely studied yet.

In this paper, we consider a *multi-source* discrete-time relaying system with *stochastic arrivals* where the sources independently generate different types of status update packets. The packets are delivered to the destination via a buffer-aided transmitter with help of a buffer-aided relay through *unreliable (error-prone)* links under transmission capacity constraint for each link. We further consider a *long-run average resource constraint* on the average number of all transmissions (i.e., transmitter-to-relay and relay-to-destination). The considered setup can be a representative of, e.g., vehicle-to-vehicle (V2V) communications, and vehicle-to-infrastructure (V2I) communications in which the infrastructure could be roadside units (RSUs) or base stations and the communications are established with the help of another vehicle acting as a relay [15]. We focus to design an *age-optimal scheduling policy* with the sum AAoI metric under the transmission capacity and average resource constraints. Specifically, we formulate a stochastic optimization problem and then recast it as a constrained Markov decision process (CMDP) problem. By adopting the Lagrangian method, we analyze the structure of an optimal policy and propose an algorithm that finds a stationary deterministic near-optimal policy. Simulation results show the effectiveness of our algorithm compared to benchmark

algorithms.

Our relaying system as a *two-hop* network is an extension of the recent work [16], where the authors provided scheduling algorithms for minimizing AoI in a *one-hop buffer-free* network with stochastic arrivals and an *error-free* link, with no resource constraint. In contrast, in our two-hop buffer-aided network, all communication links are error-prone, and we consider an average resource constraint. The most related work to our paper is [8], where the authors studied the AoI minimization problem in a single-source relaying system with the generate-at-will model under a resource constraint on the average number of forwarding transmissions at the relay. Different to [8], we consider a multi-source setup with stochastic arrivals which is a challenging generalization of the model adopted in [8]. In addition, we consider a global resource constraint on the average number of all transmissions in the system.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

We consider a status update system consisting of two independent sources<sup>1</sup>, a buffer-aided transmitter (Tx), a buffer-aided<sup>2</sup> relay (R), and a destination (D), as depicted in Fig. 1. We assume that Tx sends status update packets to D via R and there is no direct communication between Tx and D. Further, we assume that the size of the buffer in Tx and R is one packet per source. We assume that each transmission takes one slot duration.

Time is slotted and  $t = 0, 1, 2, 3, \dots$ , indicates the slot index. The sources, indexed<sup>3</sup> by  $i \in \{1, 2\}$ , generate status update packets independently according to the Bernoulli distribution with parameter  $\mu_i$ , where the packets arrive at Tx at the beginning of slots. We assume that the old packets in the waiting buffers are replaced with the new ones from the same source, which is done at the beginning of each slot. Let  $\Lambda_i[t]$  be a binary indicator that shows whether a packet from source  $i$  arrives at Tx in slot  $t$ , where  $\Lambda_i[t] = 1$  denotes that a packet arrived; otherwise,  $\Lambda_i[t] = 0$ . Thus,  $\Pr\{\Lambda_i[t] = 1\} = \mu_i$ .

• **Wireless Channels:** We assume an unreliable channel with transmission success probability  $p$  and  $q$  for the Tx-R link and R-D link, respectively. Let  $\rho[t]$  (resp.  $\rho'[t]$ ) be a binary indicator that represents the success of a packet transmission at the Tx-R link (resp. the R-D link) in slot  $t$ . In specific, if  $\rho[t] = 1$  (resp.  $\rho'[t] = 1$ ), the transmitted packet in slot  $t$  from Tx (resp. R) is successfully received by R (resp. D); otherwise,  $\rho[t] = 0$  (resp.  $\rho'[t] = 0$ ). Thus,  $\Pr\{\rho[t] = 1\} = p$  and  $\Pr\{\rho'[t] = 1\} = q$ . We assume that the perfect feedback (delay-free and error-free) is available at each link.

<sup>1</sup>We consider two sources for simplicity of presentation. The formulations and analyses can be extended for more than two sources. However, the complexity of computation, especially in numerical analysis, increases exponentially with the number of sources.

<sup>2</sup>Note that the buffers at Tx and R enable a re-transmission mechanism that can improve the AoI, in particular, when the arrival rate and transmission success probability are small. We assume that all re-transmissions have the same transmission success probability.

<sup>3</sup>Herein,  $i$  denotes the index of sources and it always varies from 1 to 2.

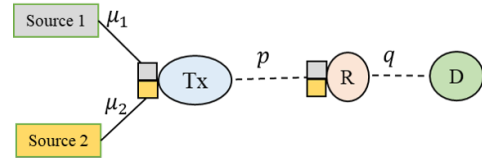


Fig. 1: A multi-source relaying status update system in which different status updates arrive at random time slots at Tx which then sends the packets to D via R. For example, in a vehicular setting, sources can be considered as different status information (e.g., speed and location), Tx can be one vehicle's transmitter called on-board unit [15], R can be another vehicle, and D can be an RSU or a vehicle.

• **Decision Variables:** We assume that, in each slot, at most one source can be scheduled per link and the transmissions are over orthogonal channels. Let  $\alpha[t] \in \{0, 1, 2\}$  denote the (transmission) decision of Tx in slot  $t$ , where  $\alpha[t] = i$  means that Tx transmits the packet of source  $i$  to R, and  $\alpha[t] = 0$  means Tx stays idle. Similarly, let  $\beta[t] \in \{0, 1, 2\}$  denote the (transmission) decision of R in slot  $t$ , where  $\beta[t] = i$  means that R forwards the packet of source  $i$  to D, and  $\beta[t] = 0$  means that R stays idle. We assume that there is a centralized controller that decides what Tx and R does during each slot.

• **AoI:** Let  $\theta_i[t]$ ,  $\delta_i[t]$ , and  $\Delta_i[t]$  be the AoI of source  $i$  at Tx, R, and D, in slot  $t$ , respectively. The evolution of these AoIs are given by

$$\begin{aligned} \theta_i[t+1] &= \begin{cases} 0, & \text{if } \Lambda_i[t+1] = 1, \\ \theta_i[t] + 1, & \text{otherwise} \end{cases}, \\ \delta_i[t+1] &= \begin{cases} \theta_i[t] + 1 & \text{if } \alpha[t] = i, \rho[t] = 1, \\ \delta_i[t] + 1, & \text{otherwise} \end{cases}, \\ \Delta_i[t+1] &= \begin{cases} \delta_i[t] + 1, & \text{if } \beta[t] = i, \rho'[t] = 1, \\ \Delta_i[t] + 1, & \text{otherwise} \end{cases}. \end{aligned}$$

### B. Problem Formulation

Let  $\mathcal{D} = \{(\alpha[t], \beta[t]) \mid \alpha[t], \beta[t] \in \{0, 1, 2\}\}$ ,  $t = 1, 2, 3, \dots$ , be a sequence of decision variables. We intend to optimize the decision variables in order to minimize the time average sum of AoIs at D, satisfying: i) the transmission capacity constraint per link at every slot and ii) a long-run average resource constraint. Let  $\bar{\Delta}(\mathcal{D})$  and  $\bar{c}(\mathcal{D})$  denote the long-run expected (time) average sum of AoIs at D and the average number of transmissions in the system, respectively, for given  $\mathcal{D}$ , which are defined as

$$\begin{aligned} \bar{\Delta}(\mathcal{D}) &\triangleq \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=1}^T \Delta_1[t] + \Delta_2[t] \right], \\ \bar{c}(\mathcal{D}) &\triangleq \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=1}^T \mathbb{1}_{\{\alpha[t] \neq 0\}} + \mathbb{1}_{\{\beta[t] \neq 0\}} \right], \end{aligned}$$

where  $\mathbb{1}_{\{\cdot\}}$  is an indicator function which equals to 1 when the condition in  $\{\cdot\}$  holds, and  $\mathbb{E}\{\cdot\}$  is the expectation with respect to the system randomness (i.e., channel reliability and packet arrival processes) and the policy. By these definitions, our aim is to solve the following stochastic optimization problem

$$\underset{\mathcal{D}}{\text{minimize}} \quad \bar{\Delta}(\mathcal{D}) \quad (1a)$$

$$\text{subject to } \bar{c}(\mathcal{D}) \leq \Gamma_{\max}, \quad (1b)$$

where the real value  $\Gamma_{\max} \in (0, 2]$  is the maximum allowable average number of transmissions in the system. Constraint (1b) is an average resource constraint that reflects the resource limitation in the normalized form, i.e., normalized to the power usage per each transmission. Thus the constraint can represent a limitation on the total average consumed power.

### III. CMDP FORMULATION AND LAGRANGE RELAXATION

In this section, we attain to solve main problem (1) by transforming it into a CMDP problem which is then solved by using the Lagrangian relaxation method.

#### A. CMDP Formulation

We introduce the CMDP by the following elements:

- **State:** The state of the CMDP incorporates the knowledge about the AoIs at Tx, R, and D. We define the state in slot  $t$  by  $\mathbf{s}[t] \triangleq (\theta_1[t], x_1[t], y_1[t], \theta_2[t], x_2[t], y_2[t])$ , where  $x_i[t] \triangleq \delta_i[t] - \theta_i[t]$  and  $y_i[t] \triangleq \Delta_i[t] - \delta_i[t]$  are the *relative AoIs* at R and D in slot  $t$ , respectively. Using the relative AoIs simplifies the subsequent analysis and derivations. The intuition is that the amount of the change of the AoI at D for source  $i$ , from slot  $t$  to the next slot  $t+1$  is proportional to  $y_i[t]$ , which is shown in the following equation

$$\Delta_i[t+1] - \Delta_i[t] = 1 - \mathbb{1}_{\{\beta[t]=i, \rho'[t]=1\}} \underbrace{(\Delta_i[t] - \delta_i[t])}_{y_i[t]}.$$

Similarly, the amount of the change of the AoI at R for source  $i$  is proportional to  $x_i[t]$ . Moreover, we denote the state space by  $\mathcal{S}$  which includes all possible states. Note that  $\mathcal{S}$  is a *countable infinite* set due to the AoIs being potentially unbounded.

- **Action:** We define the action taken in slot  $t$  by  $\mathbf{a}[t] = (\alpha[t], \beta[t])$ , where  $\alpha[t], \beta[t] \in \{0, 1, 2\}$ . Actions are determined by a policy, denoted by  $\pi$ , which is a rule that generates these actions by observing the current state (i.e., Markovian policies), i.e., a policy is a mapping from states to actions, potentially with a probability distribution. Moreover, let  $\mathcal{A}$  denote the action space.

- **Cost Functions:** The (immediate) cost functions include: 1) the AoI cost, and 2) the transmission cost. The AoI cost of each slot  $t$  is the sum AoIs at D given by  $C(\mathbf{s}[t]) = \sum_i \theta_i[t] + x_i[t] + y_i[t]$ . The transmission cost of slot  $t$ , for action  $\mathbf{a}[t]$  is defined by  $D(\mathbf{a}[t]) = \mathbb{1}_{\{\alpha[t] \neq 0\}} + \mathbb{1}_{\{\beta[t] \neq 0\}}$ .

- **State Transition Probabilities:** For any two states  $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ ,  $\mathcal{P}_{\mathbf{ss}'}$  is the state transition probability that gives the probability of moving to state  $\mathbf{s}'$  (next state) from state  $\mathbf{s}$  (current state) under taking an action  $\mathbf{a} = (\alpha, \beta)$ . Mathematically,  $\mathcal{P}_{\mathbf{ss}'}$  is given as  $\mathcal{P}_{\mathbf{ss}'}(\mathbf{a}) = \prod_i \Pr\{\mathbf{s}'_i | \mathbf{s}_i, \mathbf{a}\}$ , where  $\mathbf{s}'_i = (\theta'_i, x'_i, y'_i)$ ,  $\mathbf{s}_i = (\theta_i, x_i, y_i)$  are state vectors associated to source  $i$ , and  $\Pr\{\mathbf{s}'_i | \mathbf{s}_i, \mathbf{a}\}$  is given by (2) shown on the top of next page. Note that  $\mathbf{s}$  is arranged as  $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2)$ .

By these definitions, the objective of CMDP, denoted by  $J(\pi; \mathbf{s}[0])$ , which is the expected average sum AoIs at D, is given by

$$J(\pi; \mathbf{s}[0]) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=1}^T C(\mathbf{s}[t]) \mid \mathbf{s}[0] \right],$$

where  $\mathbf{s}[0]$  is the initial state. Similarly, the constraint function of CMDP, denoted by  $\bar{D}(\pi; \mathbf{s}[0])$ , which is the expected average transmission cost in the system, is given by

$$\bar{D}(\pi; \mathbf{s}[0]) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=1}^T D(\mathbf{a}[t]) \mid \mathbf{s}[0] \right].$$

Now, problem (1) is transformed into the following CMDP problem

$$\underset{\pi}{\text{minimize}} \quad J(\pi; \mathbf{s}[0]) \quad (3a)$$

$$\text{subject to} \quad \bar{D}(\pi; \mathbf{s}[0]) \leq \Gamma_{\max}. \quad (3b)$$

The optimal value of problem (3) for a given  $\mathbf{s}[0]$  is denoted by  $J^*(\mathbf{s}[0])$ .

#### B. Lagrange Relaxation Method

We transform problem (3) into an unconstrained average cost MDP (or simply MDP) leveraging the Lagrangian relaxation method. By this method, for a given Lagrange multiplier  $\lambda \geq 0$ , the Lagrangian, acting as the objective function for the MDP problem, is given by  $\mathcal{L}(\pi, \lambda; \mathbf{s}[0]) \triangleq \limsup_{T \rightarrow \infty} \frac{1}{T} \left( \mathbb{E} \left[ \sum_{t=1}^T C(\mathbf{s}[t]) + \lambda D(\mathbf{a}[t]) \mid \mathbf{s}[0] \right] \right)$ . Note that the term  $-\lambda \Gamma_{\max}$  is omitted from the Lagrangian because of being constant with respect to the policy for a given  $\lambda$ . An optimal policy of the MDP, for fixed  $\lambda$ , denoted by  $\pi_\lambda^*$  and called *MDP-optimal policy*, is a solution of the following MDP problem:

$$\underset{\pi_\lambda}{\text{minimize}} \quad \mathcal{L}(\pi_\lambda, \lambda; \mathbf{s}[0]). \quad (4)$$

By checking the growth condition [17, Eq. 11.20], the optimal value of the CMDP problem (3),  $J^*(\mathbf{s}[0])$ , and the optimal value of the MDP problem (4), denoted by  $\mathcal{L}^*(\lambda; \mathbf{s}[0])$ , ensures the following:

$$J^*(\mathbf{s}[0]) = \sup_{\lambda \geq 0} \mathcal{L}^*(\lambda; \mathbf{s}[0]) - \lambda \Gamma_{\max}. \quad (5)$$

Therefore, the solution of the CMDP problem (3) can be found by an algorithm that iteratively executes the following two steps: 1) find an optimal policy of the MDP problem (4) with fixed  $\lambda$ , i.e.,  $\pi_\lambda^*$ , and 2) update  $\lambda$  to a direction that aims to obtain  $J^*(\mathbf{s}[0])$  according to (5).

In the next section, we focus to find an MDP-optimal policy  $\pi_\lambda^*$  and estimating the optimal Lagrange multiplier.

### IV. OPTIMAL POLICY OF THE CMDP PROBLEM

In the following, we focus on obtaining an MDP-optimal policy and estimating the optimal Lagrange multiplier, respectively, in Sec. IV-A and Sec. IV-B.

#### A. Optimal Policy of the MDP Problem

We elaborate on the structure of MDP-optimal policy for the case with error-free links by the following theorem.

**Theorem 1.** *For given  $\lambda$  and error-free links, any MDP-optimal policy of problem (4) has a switching-type structure<sup>4</sup>*

<sup>4</sup>A switching-type structure means that if a policy takes action  $\beta = i$  at state  $\mathbf{s}$ , then it takes the same action at all states  $\mathbf{s} + z\mathbf{e}_{i+4}$ , for all  $z \in \mathbb{N}$ , where  $\mathbf{e}_{i+4}$  is a vector in  $\mathbb{B}^6$  in which the  $(i+4)$ -th element is 1 and the others are 0, where  $\mathbb{B}$  denotes the field of binary numbers.

$$\Pr\{s'_i | s_i = (\theta_i, x_i, y_i), \mathbf{a} = (\alpha, \beta)\} = \begin{cases} \mu_i pq, & \alpha = i, \beta = i; \theta'_i = 0, x'_i = \theta_i + 1, y'_i = x_i, \\ \mu_i(1-p)q, & \alpha = i, \beta = i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = 0, \\ \mu_i p(1-q), & \alpha = i, \beta = i; \theta'_i = 0, x'_i = \theta_i + 1, y'_i = y_i + x_i, \\ \mu_i(1-p)(1-q), & \alpha = i, \beta = i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = y_i, \\ (1-\mu_i)pq, & \alpha = i, \beta = i; \theta'_i = \theta_i + 1, x'_i = 0, y'_i = x_i, \\ (1-\mu_i)(1-p)q, & \alpha = i, \beta = i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = 0, \\ (1-\mu_i)p(1-q), & \alpha = i, \beta = i; \theta'_i = \theta_i + 1, x'_i = 0, y'_i = y_i + x_i, \\ (1-p)(1-q)(1-\mu_i), & \alpha = i, \beta = i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = y_i, \\ \mu_i p, & \alpha = i, \beta \neq i; \theta'_i = 0, x'_i = \theta_i + 1, y'_i = y_i + x_i, \\ \mu_i(1-p), & \alpha = i, \beta \neq i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = y_i, \\ (1-\mu_i)p, & \alpha = i, \beta \neq i; \theta'_i = \theta_i + 1, x'_i = 0, y'_i = y_i + x_i, \\ (1-\mu_i)(1-p), & \alpha = i, \beta \neq i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = y_i, \\ \mu_i q, & \alpha \neq i, \beta = i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = 0, \\ \mu_i(1-q), & \alpha \neq i, \beta = i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = y_i, \\ (1-\mu_i)q, & \alpha \neq i, \beta = i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = 0, \\ (1-\mu_i)(1-q), & \alpha \neq i, \beta = i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = y_i, \\ \mu_i, & \alpha \neq i, \beta \neq i; \theta'_i = 0, x'_i = x_i + \theta_i + 1, y'_i = y_i, \\ 1 - \mu_i, & \alpha \neq i, \beta \neq i; \theta'_i = \theta_i + 1, x'_i = x_i, y'_i = y_i, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

for  $\beta$  with respect to  $\mathbf{y} = (y_1, y_2)$ .

*Proof.* The proof will be provided in the extended version.  $\square$

Value iteration is a classical method to handle MDP problems, but cannot be applied for the infinite state spaces. To circumvent this problem, we use the state truncation method and approximation analysis [17, Ch. 16], [18].

1) *State Truncation and Approximated MDP:* We truncate  $\mathcal{S}$  into a finite state space  $\mathcal{S}^{(N)}$  which is parameterized by an integer  $N$ . To this end, whenever the AoIs exceed  $N$ , we set their values to  $N$ . The corresponding MDP is called the truncated or approximated MDP. The state transition probabilities of the truncated MDP,  $\mathcal{P}_{ss'}^{(N)}(\mathbf{a})$ , are obtained by Eq. (2) under the following changes. In all equations of Eq. (2), we replace  $\theta_i + 1$ ,  $x_i$ , and  $y_i$ , respectively, by  $\theta_i + 1 \rightarrow [\theta_i + 1]_N^+$ ,  $x_i \rightarrow [x_i + \theta_i + 1]_N^+ - [\theta_i + 1]_N^+$  and  $y_i \rightarrow [y_i + x_i + \theta_i + 1]_N^+ - [x_i + \theta_i + 1]_N^+$ , where  $[z]_N^+$  equals to  $z$  if  $z \leq N$ ; otherwise, it equals to  $N$ . In general, there is no guarantee that the truncated MDP converges to the original one [18, p. 276]. By [16, Theorem 9], for large enough  $N$ , the truncated MDP converges to the original MDP.

2) *Finding Optimal Policy of the Truncated MDP:* Now, the Relative Value Iteration (RVI) algorithm can be exploited to solve the truncated MDP for a given  $\lambda$  and parameter  $N$ . In specific, RVI is an iterative algorithm with the following value iteration equation with iteration index  $n$ :

$$V_{n+1}^{(N)}(\mathbf{s}) = \min_{\mathbf{a}} \left\{ L(\mathbf{s}, \mathbf{a}; \lambda) + \mathbb{E}[h_n^{(N)}(\mathbf{s}')] \right\},$$

where  $L(\mathbf{s}, \mathbf{a}; \lambda) = C(\mathbf{s}) + \lambda D(\mathbf{a})$  and  $h_n^{(N)}(\mathbf{s}') \triangleq V_n^{(N)}(\mathbf{s}') - V_n^{(N)}(\mathbf{s}_{\text{ref}})$  is the relative value function,  $\mathbf{s}_{\text{ref}}$  is a reference state, and  $V_0^{(N)}(\mathbf{s}_{\text{ref}}) = 0$ . Moreover,  $\mathbb{E}[h_n^{(N)}(\mathbf{s}')] = \sum_{s' \in \mathcal{S}^{(N)}} \mathcal{P}_{ss'}^{(N)}(\mathbf{a}) h_n^{(N)}(\mathbf{s}')$ . The details of RVI are provided in

Alg. 1 (see Steps 3-11), where  $\varepsilon$  is a RVI termination criterion.

For a given  $N$  and  $\lambda$ , the RVI algorithm give an optimal policy for the truncated MDP after a finite number of iterations. This is because the truncated state MDP is unichain and by [19, Theorem 8.6.6] the optimality of RVI is guaranteed.

### B. Estimating Optimal Lagrange Multiplier

By [20, Lemma 3.4], for  $\lambda > 0$ ,  $J(\pi_{\lambda}^*)$  and  $\mathcal{L}(\pi_{\lambda}^*; \lambda)$  ( $\mathbf{s}[0]$  is omitted) are increasing in  $\lambda$  and  $\bar{D}(\pi_{\lambda}^*)$  is decreasing in  $\lambda$ . Therefore, the optimal Lagrange multiplier  $\lambda^*$  is given by

$$\lambda^* = \inf\{\lambda > 0 : \bar{D}(\pi_{\lambda}^*) \leq \Gamma_{\max}\}. \quad (6)$$

As a result, if we find  $\pi_{\lambda^*}^*$  such that  $\bar{D}(\pi_{\lambda^*}^*) = \Gamma_{\max}$ , then  $\pi_{\lambda^*}^*$  is an optimal policy for the CMDP problem (3). Exploiting the monotonicity of  $\bar{D}(\pi_{\lambda}^*)$  with respect to  $\lambda$ , we adopt the bisection search [21], [22] to find  $\lambda^*$ . Details are stated in Alg. 1, where we initialize  $\lambda^- = 0$  and  $\lambda^+$  as a large positive real number, and  $\zeta$  is a sufficiently small positive real number for the bisection termination criterion.

Importantly, there is no guarantee (roughly, no possibility), even for *any arbitrarily small*  $\zeta$ , that  $\pi_{\lambda_{\text{bis}}}^*$  obtained by Alg. 1 would ensure that  $\bar{D}(\pi_{\lambda_{\text{bis}}}^*) = \Gamma_{\max}$ . This is,  $\pi_{\lambda_{\text{bis}}}^*$  (resp.  $J(\pi_{\lambda_{\text{bis}}}^*)$ ) is not necessarily an optimal policy (resp. the optimal value) for the CMDP problem (3). Following [23], by exploiting a mixing policy<sup>5</sup> which is mixture of  $\pi_{\lambda^-}^*$  (infeasible policy) and  $\pi_{\lambda^+}^*$  (feasible policy) with mixing factor  $\eta$ , the value of CMDP problem (3),  $J_{\text{mix}}$ , is given by

$$J_{\text{mix}} = \eta J(\pi_{\lambda^+}^*) + (1 - \eta) J(\pi_{\lambda^-}^*), \quad (7)$$

where

$$\eta = \frac{\Gamma_{\max} - \bar{D}(\pi_{\lambda^-}^*)}{\bar{D}(\pi_{\lambda^+}^*) - \bar{D}(\pi_{\lambda^-}^*)}. \quad (8)$$

<sup>5</sup>In a mixing policy, randomization (between two deterministic policies) occurs exactly *once* before starting to operate the system.

We remark that although  $J_{\text{mix}}$  is less than  $J(\pi_{\lambda^+}^*)$ , there is no guarantee that  $J_{\text{mix}}$  is the optimal value of the CMDP problem (3). This is because there are no guarantees that  $\pi_{\lambda^-}^*$  and  $\pi_{\lambda^+}^*$  are converged policies<sup>6</sup>, i.e., as  $\lambda^-$  increases to  $\lambda^*$ , the corresponding  $\pi_{\lambda^-}^*$  does not change (similarly for  $\lambda^+$  as it decreases to  $\lambda^*$ ) [20], [24]. However,  $J_{\text{mix}}$  serves as a benchmark to evaluate the performance of any policy such as  $\pi_{\lambda^+}^*$ .

Finally, we note that the mixing policy is not stationary and ergodic. Accordingly, one can select  $\pi_{\lambda^+}^*$  as a feasible solution for the CMDP problem (3) which is a stationary deterministic policy and is more desirable in practice. As explained before, policy  $\pi_{\lambda^+}^*$  is not guaranteed to be an optimal policy for the CMDP problem (3), while it gives acceptable performance for small enough  $\zeta$ , as shown in the numerical analysis.

---

**Algorithm 1:** Policy design for the CMDP problem (3) via RVI and bisection search

---

```

Input:  $N, \zeta, \varepsilon, \lambda^+, \lambda^-, \Gamma_{\max}, \mu_1, \mu_2, p, q$ 
1 while  $|\lambda^+ - \lambda^-| \geq \zeta$  do
2    $\lambda_{\text{bis}} = \frac{\lambda^+ + \lambda^-}{2}$ ;
3   For all  $\mathbf{s} \in \mathcal{S}^{(N)}$ , set  $V^{(N)}(\mathbf{s}) = 0, h^{(N)}(\mathbf{s}) = 0, h_{\text{old}}^{(N)}(\mathbf{s}) = 1$ ;
4   while  $\max_{\mathbf{s} \in \mathcal{S}^{(N)}} |h^{(N)}(\mathbf{s}) - h_{\text{old}}^{(N)}(\mathbf{s})| > \varepsilon$  do
5     foreach  $\mathbf{s} \in \mathcal{S}^{(N)}$  do
6        $\mathbf{a}^* \leftarrow \arg \min_{\mathbf{a}} L(\mathbf{s}, \mathbf{a}; \lambda_{\text{bis}}) + \mathbb{E}\{h^{(N)}(\mathbf{s}')\}$ ;
7        $V^{(N)}(\mathbf{s}) \leftarrow L(\mathbf{s}, \mathbf{a}^*; \lambda_{\text{bis}}) + \mathbb{E}\{h^{(N)}(\mathbf{s}')\}$ ;
8        $h_{\text{tmp}}^{(N)}(\mathbf{s}) \leftarrow V^{(N)}(\mathbf{s}) - V^{(N)}(\mathbf{s}_{\text{ref}})$ ;
9     end
10     $h_{\text{old}}^{(N)}(\mathbf{s}) \leftarrow h^{(N)}(\mathbf{s}), h^{(N)}(\mathbf{s}) \leftarrow h_{\text{tmp}}^{(N)}(\mathbf{s})$ ;
11  end
12  Compute  $\bar{D}(\pi_{\lambda_{\text{bis}}}^*)$ ;
13  if  $\bar{D}(\pi_{\lambda_{\text{bis}}}^*) \geq \Gamma_{\max}$  then
14     $\lambda^- \leftarrow \lambda_{\text{bis}}$ ;
15  else
16     $\lambda^+ \leftarrow \lambda_{\text{bis}}$ ;
17  end
18 end
19 Compute  $\pi_{\lambda^+}^*$  and  $\pi_{\lambda^-}^*$ ;
20 Compute  $\bar{D}(\pi_{\lambda^+}^*)$  and  $\bar{D}(\pi_{\lambda^-}^*)$ , and then  $\eta^*$  by (8);
21 Compute  $J(\pi_{\lambda^+}^*)$  and  $J(\pi_{\lambda^-}^*)$ ;
22 Compute  $J_{\text{mix}} = \eta J(\pi_{\lambda^+}^*) + (1 - \eta)J(\pi_{\lambda^-}^*)$ ;
Output:  $\pi_{\lambda^+}^*, J_{\text{mix}}$ 

```

---

## V. NUMERICAL RESULTS

In this section, we numerically evaluate the AoI performance of the proposed algorithm (Alg. 1) and the structure of an optimal policy. The system parameters including the values of the arrival rates,  $\mu = (\mu_1, \mu_2)$ , the reliabilities of the links in the system,  $p$  and  $q$ , and the allowable average number of transmissions,  $\Gamma_{\max}$ , are stated in the label or legend of each figure. For Alg. 1, we set  $N = 7, \zeta = 0.01$ , and  $\varepsilon = 0.001$ .

Fig. 2 shows an example of the structure of the policy for the decision at R,  $\beta$ , with respect to the relative AoIs at D,  $y_i$ , for state  $\mathbf{s} = (1, 2, 0, 1, y_1, y_2)$ . This figure verifies Theorem 1, and unveils that R schedules the available packet of the source with a higher relative AoI at D.

Fig. 3 exemplifies the structure of the policy for the decision at Tx,  $\alpha$ , with respect to the relative AoIs at R,  $x_i$ , for state

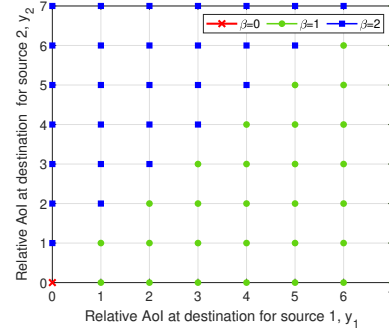


Fig. 2: An illustration of the switching-type structure for state  $\mathbf{s} = (1, 2, 0, 1, y_1, y_2)$  with respect to  $\beta$ , for  $p = 0.8, q = 0.7, \Gamma_{\max} = 1.6$ , and  $\mu = (0.6, 0.9)$ .

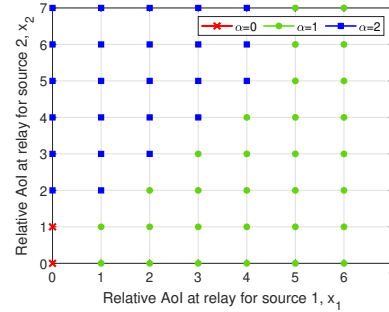


Fig. 3: An illustration of the switching-type structure for state  $\mathbf{s} = (1, 1, x_1, x_2, 4, 4)$  with respect to  $\alpha$ , for  $p = 0.8, q = 0.7$ , and  $\Gamma_{\max} = 1.6, \mu = (0.6, 0.9)$ .

$\mathbf{s} = (1, 1, x_1, x_2, 4, 4)$ . Having  $\alpha = 0$  at  $(x_1 = 0, x_2 = 1)$  implies that transmission does not occur at every state due to the resource budget. We can conclude from this figure that for fixed  $y_1$  and  $y_2$ , Tx will give a higher priority to schedule the source that has a lower status update rate. This is because the waiting time at D for receiving new updates from that source becomes large as a consequence of infrequent packet arrivals, which becomes a bottleneck in reducing the AoI.

Fig. 4 depicts the average sum AoI at D (sum AAoI) with respect to the allowable average number of transmissions in the system (resource budget),  $\Gamma_{\max}$ , for different arrival rates and reliability of the links, obtained by averaging over 100,000 time slots. In this figure, “Mix.” refers to  $J_{\text{mix}}$  obtained by (7), and “Deter.” stands for  $J(\pi_{\lambda^+}^*)$ . For benchmarking, we consider a “Lower bound” scheme where we eliminate the average resource constraint by setting  $\Gamma_{\max} = 2$  and realize the *generate-at-will* model by setting the arrival rates as  $\mu_1 = \mu_2 = 1$ . With these setting, our system becomes equivalent to the one studied in [6], and we use the greedy-based optimal policy derived in [6]. As another benchmark, we design a “Greedy” policy, where the transmission is allowed in slot  $t$  when  $\bar{D}_t \leq \Gamma_{\max}$ , where  $\bar{D}_t$  denotes the average number of transmissions until  $t$ , and the decision criteria are the relative AoI at R for  $\alpha$  and the relative AoI at D for  $\beta$ .

First, Fig. 4 shows that the deterministic policy,  $\pi_{\lambda^+}^*$ , achieves a near-optimal performance. Fig. 4 illustrates that the sum AAoI dramatically increases when the resource budget  $\Gamma_{\max}$  is decreased, and the sum AAoI values become large

<sup>6</sup>Here, convergence is point-wise, see [20, Definition 3.6].

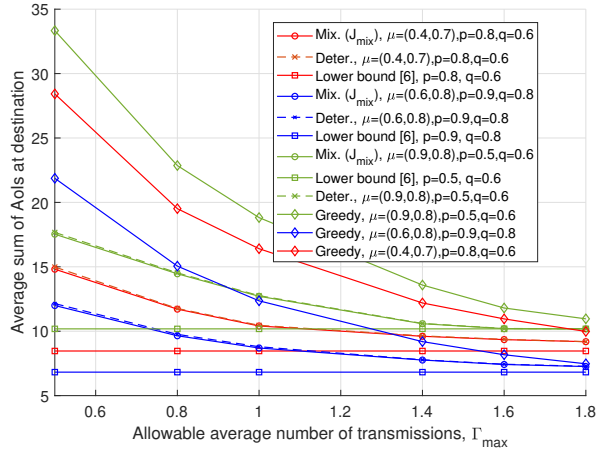


Fig. 4: Average sum AoIs at D versus  $\Gamma_{\max}$ , for different arrival rates  $\mu = (\mu_1, \mu_2)$ , and reliability of channels,  $p$  and  $q$ .

when channel reliabilities are reduced. Moreover, Fig. 4 exhibit that the sum AAOI converges to the lower bound when the arrival rates of updates are high and the resource budget is large. For that case, we infer that an optimal policy has a greedy behavior where the maximum of the relative AoIs at R is the greedy criterion for the decision of Tx, and the maximum of the relative AoIs at D is a greedy criterion for the decision of R. We observe that the optimality gap for the Greedy policy is large when the resource budget is small. This clearly emphasizes that it is pivotal to take into account resource limitations in age-optimal policy design. Namely, it may be possible to find some algorithms that minimize the sum AAOI, but they are not necessarily resource-efficient ones.

## VI. CONCLUSION

We studied the sum AAOI minimization scheduling problem in a multi-source relaying system with stochastic arrivals and unreliable communication channels, under per-slot transmission capacity constraints per link and a long-run average resource constraint. To this end, we formulated a stochastic optimization problem and recast it as a CMDP problem. We analyzed the structure of an optimal policy and proposed an algorithm that obtains a stationary deterministic near-optimal policy which is easy to implement. According to the simulation results, our proposed algorithm significantly reduces the sum AAOI compared with the non-trivial greedy-based benchmark algorithm. We concluded that when the average resource budget is large and the arrival rates of status updates are high, an optimal policy has a greedy behavior.

## VII. ACKNOWLEDGMENTS

This research has been financially supported by the Infotech Oulu, the Academy of Finland (grant 323698), and Academy of Finland 6Genesis Flagship (grant 318927). The work of M. Leinonen has also been financially supported in part by the Academy of Finland (grant 319485).

- [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] E. Altman, R. El-Azouzi, D. S. Menasche, and Y. Xu, "Forever young: Aging control for smartphones in hybrid networks," CoRR, Sep. 2010.
- [3] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE Int. Conf. on Computer Commun.*, pp. 2731–2735, Orlando, FL, USA, Mar. 2012.
- [4] M. Moltafet, M. Leinonen, and M. Codreanu, "On the age of information in multi-source queueing models," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5003–5017, Aug. 2020.
- [5] 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.
- [6] J. Song, D. Gunduz, and W. Choi, "Optimal scheduling policy for minimizing age of information with a relay," *arXiv preprint arXiv:2009.02716*, Sep. 2020.
- [7] M. Hatami, M. Leinonen, and M. Codreanu, "AoI minimization in status update control with energy harvesting sensors," *arXiv preprint arXiv:2009.04224*, Sep. 2020.
- [8] Y. Gu, Q. Wang, H. Chen, Y. Li, and B. Vucetic, "Optimizing information freshness in two-hop status update systems under a resource constraint," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1380–1392, May, 2021.
- [9] M. Moltafet, M. Leinonen, M. Codreanu, and N. Pappas, "Power minimization for age of information constrained dynamic control in wireless sensor networks," *arXiv preprint arXiv:2007.05364*, Jul. 2020.
- [10] M. Moradian and A. Dadlani, "Age of information in scheduled wireless relay networks," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, pp. 1–6, Seoul, Korea (South), Mar. 2020.
- [11] B. Li, H. Chen, N. Pappas, and Y. Li, "Optimizing information freshness in two-way relay networks," in *Proc. IEEE/CIC Int. Conf. on Commun. in China (ICCC)*, pp. 893–898, Chongqing, China, Aug. 2020.
- [12] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 576–580, Aachen, Germany, Jun. 2017.
- [13] A. Arafa and S. Ulukus, "Timely updates in energy harvesting two-hop networks: Offline and online policies," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4017–4030, Aug. 2019.
- [14] B. Li, Q. Wang, H. Chen, Y. Zhou, and Y. Li, "Optimizing information freshness for cooperative IoT systems with stochastic arrivals," *IEEE Internet Things J*, Early Access, 2021.
- [15] Z. Su, Y. Hui, T. H. Luan, Q. Liu, and R. Xing, *The Next Generation Vehicular Networks, Modeling, Algorithm and Applications*. Springer, 2020.
- [16] 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.
- [17] E. Altman, *Constrained Markov Decision Processes*. volume 7. CRC Press, 1999.
- [18] L. I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*. volume 504. John Wiley & Sons, 2009.
- [19] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. The MIT Press, 1994.
- [20] L. I. Sennott, "Constrained average cost Markov decision chains," *Probab. Eng. Inf. Sci.*, vol. 7, no. 1, pp. 69–83, Jan. 1993.
- [21] H. Huang, D. Qiao, and M. C. Gursoy, "Age-energy tradeoff in fading channels with packet-based transmissions," in *Proc. IEEE Int. Conf. on Computer Commun.*, pp. 323–328, Toronto, ON, Canada, Jul., 2020.
- [22] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: an enabler of semantics-empowered communication," *arXiv preprint arXiv:2012.13214*, Dec. 2020.
- [23] D.-J. Ma, M. Makowski, and A. Shwartz, "Estimation and optimal control for constrained Markov chains," in *Proceedings of the 25th IEEE Conf. on Decision and Control*, pp. 994–999, 1987.
- [24] K. Beutler, F.J. Ross, "Optimal policies for controlled Markov chains with a constraint," *Mathematical Methods of Operations Research*, vol. 112, pp. 236–252., 1985.