

# Optimizing the Trade-off Between Throughput and PAoI Outage Exponents

Tai-Chun Yeh and Yu-Pin Hsu

Department of Communication Engineering

National Taipei University

a88700337@gmail.com, yupinhsu@mail.ntpu.edu.tw

**Abstract**—This paper investigates the trade-off between throughput and peak age of information (PAoI) outage probability in a multi-sensor information collection system. Each sensor monitors a physical process, periodically samples its status, and transmits the updates to a central access point over a shared radio resource. The trade-off arises from the interplay between each sensor's sampling frequency and the allocation of the shared resource. To optimize this trade-off, we formulate a joint optimization problem for each sensor's sampling delay and resource allocation, aiming to minimize a weighted sum of sampling delay costs (representing a weighted sum of throughput) while satisfying PAoI outage probability exponent constraints. We derive an optimal solution and particularly propose a closed-form approximation for large-scale systems. This approximation provides an explicit expression for an approximately optimal trade-off, laying a foundation for designing resource-constrained systems in applications that demand frequent updates and also stringent statistical timeliness guarantees.

## I. INTRODUCTION

In recent years, the rapid advancement of smart systems has significantly increased the demand for timely information to ensure their efficient and effective operation. For instance, in a smart transportation system [1], vehicles are equipped with sensors (e.g., GPS or radar) that continuously monitor their surroundings. These sensors generate updates about various physical processes and transmit the updates to a central controller. By aggregating these updates, the central controller gains a real-time understanding of the dynamic environment, enabling smart decision-making. To ensure safety and reliability, the information maintained by the central controller must remain as timely as possible.

To quantify the timeliness of information, [2] introduced the concept of the age of information (AoI), which measures the time elapsed since the generation of the most recent update. A lower AoI indicates that the received update is closer to the current state. Building on this, [3] proposed the peak age of information (PAoI), which represents the maximum AoI observed just before a new update is received. PAoI is particularly valuable for systems where worst-case timeliness or adherence to specific age thresholds is critical.

Extensive research has been conducted on analyzing and minimizing average AoI/PAoI in diverse system settings. For instance, [4] derived analytical expressions for the average

AoI and PAoI in single-source systems, while [3] extended this analysis to multi-source systems. Moreover, numerous system design strategies have been proposed to minimize these metrics, e.g., scheduling [5], resource allocation [6], and sampling [7]. Comprehensive surveys of these efforts can be found in [8, 9]. However, while minimizing the average AoI/PAoI typically enhances timeliness, it does not ensure a desired AoI with high probability. For systems with strict timeliness requirements, statistical AoI/PAoI guarantees are crucial. This has motivated research into the probabilistic characterization of AoI/PAoI, e.g., [10–12].

In addition to timeliness, the total number of status updates (usually referred to as throughput) is a critical factor influencing the performance of smart systems. For example, in a smart transportation system, frequent sampling of a vehicle's location allows the central controller to predict its future position with greater accuracy. However, overly frequent sampling can lead to network congestion, causing higher queueing delay and ultimately stale updates. To address this, numerous studies have examined the trade-off between throughput and *average* AoI/PAoI, e.g., [13, 14]. Despite these efforts, the relationship between throughput and *statistical* AoI/PAoI remains unclear. This gap is particularly relevant for systems that demand frequent updates while adhering to stringent statistical AoI/PAoI guarantees, such as those employed in safety-critical applications.

To address this gap, this paper investigates the trade-off between throughput and PAoI outage probability in an information status collection system. The system consists of multiple local monitors that observe physical processes and periodically generate status updates, which are transmitted to a central controller over a shared communication resource. Since PAoI is influenced by both sampling delay and transmission delay (the latter determined by resource allocation), optimizing only one of these factors is insufficient to achieve the optimal trade-off.

The primary contribution of this paper is the joint optimal design of sampling delay and resource allocation. The objective is to minimize a total sampling delay cost while satisfying PAoI outage probability exponent constraints, providing a framework to analyze the trade-off between throughput and PAoI outage. Our main result reveals that, if the transmission time is exponentially distributed, then an approximately optimal resource allocation consists of the minimum required allocation to satisfy the PAoI outage constraints, supplemented

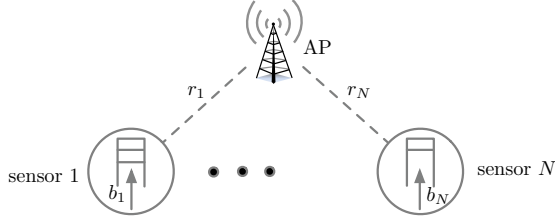


Fig. 1. An information collection system.

by a proportional distribution of the remaining resource. Moreover, through the approximation, we derive a closed-form expression that captures the relationship between each sensor's sampling delay (indicative of each sensor's throughput) and the PAoI outage probability exponents under the approximately optimal resource allocation. These findings are critical for designing systems that require frequent and timely information updates.

## II. SYSTEM OVERVIEW

### A. Network model

As illustrated in Fig. 1, we consider an information status collection system comprising  $N$  sensors and an access point (AP). Each sensor monitors a physical process and updates the AP with the latest status of the process. Specifically, each sensor samples the physical process periodically. Let  $b_i$  represent the sampling delay for sensor  $i$ , and define  $\mathbf{b} = (b_1, \dots, b_N)$ . Additionally, let  $S_{i,j}$  denote the time when sensor  $i$  generates sample  $j$ .

The generated samples are transmitted to the AP as packets using a shared radio resource. Let  $r_i$  represent the fraction of the total resource allocated to sensor  $i$ , satisfying  $\sum_{i=1}^N r_i \leq 1$ . Define  $\mathbf{r} = (r_1, \dots, r_N)$ . Given a resource allocation  $\mathbf{r}$ , the transmission time of a packet is assumed to be independent of other packets and follows the same statistical distribution for packets from the same sensor. Let  $T_i$  denote the transmission time of a packet from sensor  $i$ , where the distribution of  $T_i$  depends on the allocated resource  $r_i$  only. Each sensor maintains a queue to store packets awaiting transmission, and the packets in the queue are transmitted in a first-come first-served manner.

To evaluate the timeliness of information at the AP, we use the peak age of information (PAoI) metric proposed in [3]. PAoI quantifies the maximum staleness of information just before an update is received at the AP. Formally, let  $D_{i,j}$  represent the time when packet  $j$  from sensor  $i$  is delivered to the AP. PAoI for sensor  $i$ , denoted by  $A_{i,j}$ , upon receiving packet  $j$  is given by  $A_{i,j} = D_{i,j} - S_{i,j-1}$ . Let  $A_{i,\infty} = \lim_{j \rightarrow \infty} A_{i,j}$  denote the steady-state PAoI for sensor  $i$ . Unlike most existing studies that focused on average AoI/PAoI, this work examines outage probability of PAoI, which is crucial for applications requiring strict statistical guarantees. Specifically, the AP imposes a constraint  $\theta_i > 0$  on sensor  $i$  such that

$$\lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} \leq -\theta_i,$$

ensuring that the outage probability  $\mathbb{P}[A_{i,\infty} \geq x]$  in the steady state decays exponentially with a rate of at least  $\theta_i$ , as  $x$  increases. Here,  $\theta_i$  is referred to as the outage exponent. The higher the value of  $\theta_i$ , the more stringent the probabilistic PAoI requirement. Let  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ .

### B. Problem formulation

To provide more timely information to the AP, a shorter sampling delay is desired, which, however, increases queueing delay and consequently raises PAoI. To study this trade-off, let  $C_i$  denote the cost of delaying sampling for sensor  $i$  by one unit of time, and define  $\mathbf{C} = (C_1, \dots, C_N)$ . The goal is to minimize the total sampling delay cost  $\sum_{i=1}^N C_i b_i$  (reflecting a high weighted sum of throughput since a lower  $b_i$  corresponds to a higher sampling frequency) while ensuring the outage exponents for all sensors are satisfied.

Since PAoI depends on both sampling delay  $\mathbf{b}$  and resource allocation  $\mathbf{r}$ , the problem requires jointly designing  $\mathbf{b}$  and  $\mathbf{r}$ . Precisely, given outage exponent  $\boldsymbol{\theta}$  and sampling delay cost  $\mathbf{C}$ , we formulate the following optimization problem:

$$\min_{\mathbf{b}, \mathbf{r}} \sum_{i=1}^N C_i b_i \quad (1a)$$

$$\text{s.t. } \lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} \leq -\theta_i, \text{ for all } i = 1, \dots, N; \quad (1b)$$

$$\sum_{i=1}^N r_i \leq 1. \quad (1c)$$

Note that this formulation captures the trade-off between the total sampling delay cost and the statistical PAoI guarantees in a heterogeneous network, where different sensors may have distinct sampling delay costs or transmission times.

## III. OPTIMAL DESIGN FOR SAMPLING DELAY

In this section, we characterize the optimality structure of the sampling delay design. To this end, let  $\Lambda_i(\theta) = \ln \mathbb{E}[e^{\theta T_i}]$  represent the log moment generating function (LMGF) of transmission time  $T_i$ . From [10, Lemma 1], if the condition  $\frac{\Lambda_i(\theta_i)}{\theta_i} < b_i$  (with strict inequality) is satisfied, then the asymptotic behavior of PAoI satisfies  $\lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} \leq -\theta_i$ , ensuring that the outage exponent  $\theta_i$  is achievable. Building on this result, we further establish a stronger characterization in the following lemma. Specifically, we show that the condition  $\frac{\Lambda_i(\theta_i)}{\theta_i} \leq b_i$  is not only sufficient but also necessary for satisfying the outage constraint in Eq. (1b).

**Lemma 1.** *The condition  $\lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} \leq -\theta_i$  holds if and only if  $\frac{\Lambda_i(\theta_i)}{\theta_i} \leq b_i$ .*

*Proof (sketch):* We employ large deviation theory [15] to show that the AoI outage probability exponent can be expressed by  $\lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} = -\inf_{t > 0} \frac{1}{t} I_i(t + b)$ , where  $I_i(x) = \sup_{\gamma \in \mathbf{R}} \{\gamma x - \Lambda_i(\gamma)\}$  is the rate function corresponding to transmission time  $T_i$ . Then, leveraging the relation between the rate function  $I_i(\cdot)$  and the LMGF  $\Lambda_i(\cdot)$ , we can establish the equivalence. The detailed proof is given in Appendix A. ■

Lemma 1 is analogous to the effective bandwidth theory (e.g., see [16]) in queueing systems; however, their results cannot be directly applied here, as the evolution of PAoI differs from that of queues. Moreover, from Lemma 1, the constraint in Eq. (1b) can be equivalently replaced by  $\frac{\Lambda_i(\theta_i)}{\theta_i} \leq b_i$ . Using this substitution, the optimization problem in Eq. (1) can be reformulated as the following equivalent problem:

$$\min_{\mathbf{b}, \mathbf{r}} \sum_{i=1}^N C_i b_i \quad (2a)$$

$$\text{s.t. } \frac{\Lambda_i(\theta_i)}{\theta_i} \leq b_i, \text{ for all } i = 1, \dots, N; \quad (2b)$$

$$\sum_{i=1}^N r_i \leq 1. \quad (2c)$$

Recall that the LMGF  $\Lambda_i(\theta_i)$  of transmission time  $T_i$  depends solely on resource allocation and is independent of sampling delay. Therefore, from Eq. (2b), for a given resource allocation  $\mathbf{r}$ , the sampling delay for sensor  $i$  that minimizes the total sampling delay cost in Eq. (2a) can be determined as  $b_i = \frac{\Lambda_i(\theta_i)}{\theta_i}$ . In the next section, we will derive an optimal resource allocation, denoted by  $\mathbf{r}^*$ . Once  $\mathbf{r}^*$  is determined, the optimal sampling delay for sensor  $i$ , denoted by  $b_i^*$ , can be computed as

$$b_i^* = \frac{\Lambda_i^*(\theta_i)}{\theta_i}, \quad (3)$$

where  $\Lambda_i^*(\cdot)$  represents the LMGF of  $T_i$  under the optimal resource allocation  $\mathbf{r}_i^*$ .

#### IV. OPTIMAL DESIGN FOR RESOURCE ALLOCATION UNDER A SPECIFIC TRANSMISSION DISTRIBUTION

In this section, we determine the optimal resource allocation for a specific transmission time distribution. We model transmission time  $T_i$  as an exponential random variable with rate  $\mu_i r_i$ , where  $\mu_i$  represents the mean transmission rate per unit of resource. This model implies that the transmission capacity of a sensor scales proportionally with the amount of resource allocated to it.

For exponentially distributed transmission times, the LMGF  $\Lambda_i(\theta_i)$  can be computed as  $\Lambda_i(\theta_i) = \ln \frac{\mu_i r_i}{\mu_i r_i - \theta_i}$ . Given the relationship between  $r_i^*$  and  $b_i^*$  derived in Eq. (3), we substitute  $b_i$  in Eq. (2a) with  $b_i = \frac{1}{\theta_i} \ln \frac{\mu_i r_i}{\mu_i r_i - \theta_i}$ . This substitution transforms the optimization problem in Eq. (2) into the following equivalent form for determining the optimal resource allocation  $\mathbf{r}$ :

$$\min_{\mathbf{r}} \sum_{i=1}^N \frac{C_i}{\theta_i} \cdot \ln \frac{\mu_i r_i}{\mu_i r_i - \theta_i} \quad (4a)$$

$$\text{s.t. } \sum_{i=1}^N r_i \leq 1. \quad (4b)$$

##### A. Feasibility region of outage exponent $\theta$

We note that the objective function in Eq. (4a) is feasible only when  $\mu_i r_i > \theta_i$ . This implies that if sufficient resource  $r_i$  cannot be allocated to sensor  $i$  to ensure its transmission

rate  $\mu_i r_i$  exceeds its outage exponent  $\theta_i$ , the outage exponent guarantee cannot be achieved (as stated in Lemma 1). Moreover, if no resource allocation  $\mathbf{r}$  satisfies  $\mu_i r_i > \theta_i$  for all  $i$  and  $\sum_{i=1}^N r_i \leq 1$ , the optimization problem in Eq. (4) has no solution.

Thus, we define outage exponent  $\theta$  as *feasible* if and only if there exists a resource allocation  $\mathbf{r}$  such that  $\mu_i r_i > \theta_i$  for all  $i$  and also  $\sum_{i=1}^N r_i \leq 1$ . The following lemma provides a necessary and sufficient condition for the feasibility of  $\theta$ .

**Lemma 2.** A set  $\theta$  of outage exponents is feasible if and only if  $\sum_{i=1}^N \frac{\theta_i}{\mu_i} < 1$ .

*Proof:* First, assume that  $\sum_{i=1}^N \frac{\theta_i}{\mu_i} < 1$ . Then, there exists an  $\epsilon > 0$  such that  $\epsilon \leq \frac{1 - \sum_{i=1}^N \frac{\theta_i}{\mu_i}}{N}$ . Consider a resource allocation for sensor  $i$  as  $r_i = \frac{\theta_i}{\mu_i} + \epsilon$ . With this allocation,  $\mu_i r_i > \theta_i$  for all  $i$ . Additionally, we have

$$\begin{aligned} \sum_{i=1}^N r_i &= \sum_{i=1}^N \left( \frac{\theta_i}{\mu_i} + \epsilon \right) = \sum_{i=1}^N \frac{\theta_i}{\mu_i} + N\epsilon \\ &\leq \sum_{i=1}^N \frac{\theta_i}{\mu_i} + \left( 1 - \sum_{i=1}^N \frac{\theta_i}{\mu_i} \right) = 1. \end{aligned}$$

Thus, the outage exponent  $\theta$  is feasible.

Next, assume that  $\sum_{i=1}^N \frac{\theta_i}{\mu_i} \geq 1$ , and suppose  $\mu_i r_i > \theta_i$  for all  $i$ . Then, we have

$$\sum_{i=1}^N r_i > \sum_{i=1}^N \frac{\theta_i}{\mu_i} \geq 1,$$

which contradicts the constraint  $\sum_{i=1}^N r_i \leq 1$ . Therefore, by contraposition,  $\sum_{i=1}^N \frac{\theta_i}{\mu_i} \geq 1$  implies that the outage exponent  $\theta$  is not feasible. ■

As a result, the set  $\{\theta : \sum_{i=1}^N \frac{\theta_i}{\mu_i} < 1\}$  defines the region of all feasible outage exponents. This feasibility region is analogous to the capacity region [15] in queueing systems. In this paper, we focus on scenarios where the requested outage exponent  $\theta$  lies within the feasibility region.

##### B. Optimal solution for resource allocation

The optimal resource allocation can be determined by solving the optimization problem in Eq. (4). Since the problem in Eq. (4) is a convex optimization problem and satisfies the Slater condition for a feasible  $\theta$ , it can be solved using the Karush-Kuhn-Tucker (KKT) conditions [15].

The Lagrangian for this problem is given by

$$\mathcal{L}(\mathbf{r}, \lambda) = \sum_{i=1}^N \left( \frac{C_i}{\theta_i} \cdot \ln \frac{\mu_i r_i}{\mu_i r_i - \theta_i} \right) + \lambda \left( \sum_{i=1}^N r_i - 1 \right),$$

where  $\lambda > 0$  is a Lagrange multiplier. Differentiating the Lagrangian with respect to  $r_i$ , we obtain

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{r}, \lambda)}{\partial r_i} &= \frac{C_i}{\theta_i} \cdot \left( \frac{\mu_i r_i - \theta_i}{\mu_i r_i} \cdot \frac{\mu_i (\mu_i r_i - \theta_i) - \mu_i^2 r_i}{(\mu_i r_i - \theta_i)^2} \right) + \lambda \\ &= \frac{-C_i}{r_i (\mu_i r_i - \theta_i)} + \lambda. \end{aligned}$$

Setting the derivative to zero, we can obtain  $\mu_i \lambda r_i^2 - \theta_i \lambda r_i - C_i = 0$ . Solving this quadratic equation, the optimal resource allocation for sensor  $i$  is

$$r_i^* = \frac{\theta_i \lambda + \sqrt{(\theta_i \lambda)^2 + 4C_i \mu_i \lambda}}{2\mu_i \lambda} = \frac{\theta_i}{2\mu_i} + \frac{\theta_i}{2\mu_i} \left(1 + \frac{4C_i \mu_i}{\theta_i^2 \lambda}\right)^{\frac{1}{2}}. \quad (5)$$

The Lagrange multiplier  $\lambda$  can be determined by enforcing the constraint  $\sum_{i=1}^N r_i^* = 1$ , leading to

$$\sum_{i=1}^N \left( \frac{\theta_i}{2\mu_i} + \frac{\theta_i}{2\mu_i} \left(1 + \frac{4C_i \mu_i}{\theta_i^2 \lambda}\right)^{\frac{1}{2}} \right) = 1. \quad (6)$$

Due to the square root terms in Eq. (6), there is no closed-form solution for  $\lambda$ . However, this equation can be solved using numerical methods, such as Newton's method [17].

## V. APPROXIMATE SOLUTION

To gain insights into the optimal resource allocation in Eq. (5) and also characterize the optimal trade-off between throughput and P<sub>AoI</sub> outage exponents, this section presents an approximation. When the number  $N$  of sensors is large,  $\lambda$  becomes large, as indicated by Eq. (6). Using the approximation  $(1+x)^{\frac{1}{2}} \approx 1 + \frac{x}{2}$  when  $x$  is small, we can approximate Eq. (5) as

$$r_i^* \approx \frac{\theta_i}{2\mu_i} + \frac{\theta_i}{2\mu_i} \left(1 + \frac{1}{2} \cdot \frac{4C_i \mu_i}{\theta_i^2 \lambda}\right) = \frac{\theta_i}{\mu_i} + \frac{C_i}{\theta_i} \frac{1}{\lambda}. \quad (7)$$

According to Eq. (7), we can interpret  $\theta_i$  as the required transmission rate for sensor  $i$  to satisfy its outage exponent. Thus, the first term  $\frac{\theta_i}{\mu_i}$  in Eq. (7) corresponds to the minimum resource required for sensor  $i$  to meet its outage exponent. The AP reserves  $\sum_{i=1}^N \frac{\theta_i}{\mu_i}$  resource to satisfy all outage exponents. The remaining resource,  $1 - \sum_{i=1}^N \frac{\theta_i}{\mu_i}$ , is distributed to minimize the total sampling delay cost.

The second term in Eq. (7) shows that the remaining resource is allocated proportional to  $\frac{C_i}{\theta_i}$ . This allocation can be explained as follows: From Eq. (3), the sampling delay  $b_i^*$  is given by  $b_i^* = \frac{1}{\theta_i} \ln \frac{\mu_i r_i^*}{\mu_i r_i^* - \theta_i}$ . Allocating more resource  $r_i^*$  reduces  $b_i^*$ , lowering the total sampling delay cost. Also, a higher  $\theta_i$  increases  $b_i^*$ , as a stricter outage exponent requires less frequent sampling to avoid excessive queueing delay. Thus, the remaining resource allocation follows the principles as follows:

- A sensor  $i$  with a high sampling delay cost  $C_i$  is allocated more remaining resource to reduce its sampling delay and offset the high unit cost.
- A sensor  $i$  with a stringent outage exponent  $\theta_i$  is allocated fewer remaining resource, resulting in a longer sampling delay to meet the stricter constraint.

Next, using the condition  $\sum_{i=1}^N r_i^* = 1$ , the Lagrange multiplier  $\lambda$  in Eq. (7) can be derived, and the approximately optimal resource allocation becomes

$$r_i^* \approx \frac{\theta_i}{\mu_i} + \frac{C_i}{\theta_i} \frac{1 - \sum_{i=1}^N \frac{\theta_i}{\mu_i}}{\sum_{i=1}^N \frac{C_i}{\theta_i}}. \quad (8)$$

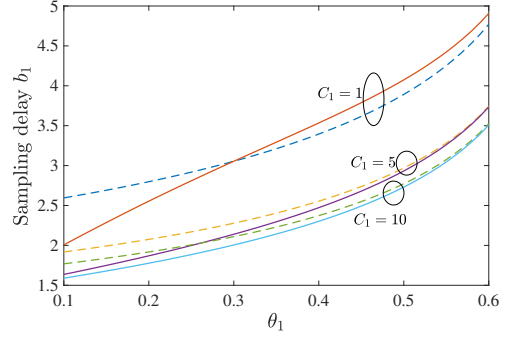


Fig. 2. Sampling delay  $b_1$  versus outage exponent  $\theta_1$ .

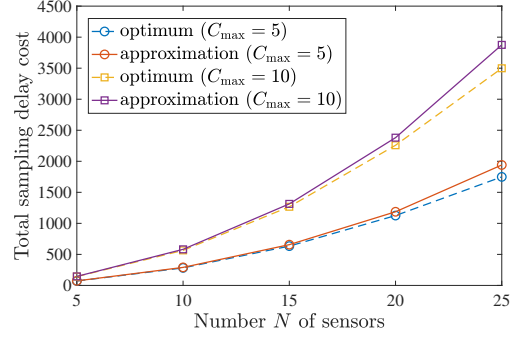


Fig. 3. The total sampling delay cost versus the number  $N$  of sensors.

Substituting this allocation into Eq. (3), the approximately optimal sampling delay becomes

$$b_i^* \approx \frac{1}{\theta_i} \ln \left( 1 + \frac{\theta_i^2}{C_i \mu_i} \frac{\sum_{i=1}^N \frac{C_i}{\theta_i}}{1 - \sum_{i=1}^N \frac{\theta_i}{\mu_i}} \right). \quad (9)$$

Our approximation in Eqs. (8) and (9) provides a computationally efficient approach to determine the resource allocation and the sampling delay, avoiding the need for complex numerical methods. Moreover, Eq. (9) characterizes an approximately optimal trade-off between sampling frequency and outage exponents across various system parameters in a heterogeneous network. The numerical analysis of this trade-off will be discussed in the next section.

## VI. NUMERICAL RESULTS

In this section, we evaluate the proposed approximate solution through numerical studies. For simplicity, we set  $\mu_i = 1$  for all  $i$ .

First, we consider a system with two sensors, where  $\theta_2 = 0.3$  and  $C_2 = 1$  are fixed. Fig. 2 illustrates the trade-off between sampling delay  $b_1$  and outage exponent  $\theta_1$  under the optimal resource allocation (dashed curves) and the approximate solution (solid curves). From Fig. 2, we observe that the approximate trade-off closely matches the optimal trade-off when  $\theta_1$  approaches its maximum feasible value (i.e., when  $\theta$  is near the boundary of the feasibility region) or when the sampling delay cost  $C_1$  becomes larger.

Second, we examine the accuracy of the approximation in terms of the total sampling delay cost (indicative of a weighted

sum of throughput), as shown in Fig. 3. Here, the outage exponents are set as follows:  $\theta_{\frac{N}{2}} = \frac{0.5}{N}$ ,  $\theta_i = \frac{0.5}{N} - 0.01(\frac{N}{2} - i)$  for  $i < \frac{N}{2}$ , and  $\theta_i = \frac{0.5}{N} + 0.01(i - \frac{N}{2})$  for  $i > \frac{N}{2}$ . This configuration ensures  $\sum_{i=1}^N \theta_i = 0.5$ , which is far from the boundary of the feasibility region. The sampling delay cost  $C_i$  is randomly selected from the range  $[1, C_{\max}]$  for all  $i$ , and each data point in the figure is averaged over 1000 random scenarios. From Fig. 3, we observe that the approximation performs closely to the optimal solution (in terms of total sampling delay cost), even for small values of  $N$ .

## VII. CONCLUSION

In this paper, we investigated an information collection system operating over a shared communication resource. We proposed a joint design of sampling and resource allocation to approximately minimize the total sampling delay cost while satisfying PAoI outage exponent constraints. While this work focuses on a feasible set of outage exponents, an interesting direction for future research is the development of admission control mechanisms for systems where the outage exponents exceed the feasibility region.

### APPENDIX A PROOF OF LEMMA 1

The proof considers a fixed sensor  $i$ . From [10], PAoI can be expressed by  $A_{i,j} = \max_{1 \leq s \leq j} \{\sum_{k=s}^j T_{i,k} - (j-1-s)b\}$ , which can be expanded as  $\max\{T_{i,j}+b, T_{i,j-1}+T_{i,j}, T_{i,j-2}+T_{i,j-1}+T_{i,j}-b, \dots\}$ . Because of i.i.d. transmission times  $T_{i,j}$  for different  $j$ , we can re-express PAoI as  $A_{i,j} = \max_{1 \leq s \leq j} \{\sum_{k=1}^s T_{i,k} - sb + 2b\}$ . In the steady state, as  $j \rightarrow \infty$ , this becomes  $A_{i,\infty} = \max_{s \geq 1} \{\sum_{k=1}^s T_{i,k} - sb + 2b\}$ .

Let  $I_i(x) = \sup_{\gamma \in \mathbf{R}} \{\gamma x - \Lambda_i(\gamma)\}$  be the rate function associated with transmission time  $T_i$ . We first derive an upper bound on the outage probability:

$$\begin{aligned} \mathbb{P}[A_{i,\infty} \geq x] &= \mathbb{P}\left[\max_{s \geq 1} \left\{\sum_{k=1}^s T_{i,k} - sb + 2b\right\} \geq x\right] \\ &\leq \sum_{s=1}^{\infty} \mathbb{P}\left[\sum_{k=1}^s T_{i,k} - sb + 2b \geq x\right] \\ &= \sum_{s=1}^{\infty} \mathbb{P}\left[\sum_{k=1}^s T_{i,k} \geq x + sb - 2b\right] \\ &\stackrel{(a)}{\leq} \sum_{s=1}^{\infty} \frac{\mathbb{E}\left[e^{\gamma \sum_{k=1}^s T_{i,k}}\right]}{e^{\gamma(x+sb-2b)}} \text{ for any } \gamma > 0 \\ &\stackrel{(b)}{=} \sum_{s=1}^{\infty} \frac{e^{s\Lambda_i(\gamma)}}{e^{\gamma(x+sb-2b)}} \\ &= e^{2\gamma b} \sum_{s=1}^{\infty} e^{-s(\gamma(\frac{x}{s}+b)-\Lambda_i(\gamma))} \\ &\stackrel{(c)}{=} e^{2\gamma b} \sum_{t \in \{\frac{1}{x}, \frac{2}{x}, \dots\}} e^{-xt(\gamma(\frac{1}{t}+b)-\Lambda_i(\gamma))}, \end{aligned}$$

where step (a) applies the Chernoff bound, step (b) uses the i.i.d. assumption for  $T_{i,k}$ 's, and step (c) defines another variable  $t = \frac{s}{x}$ . Note that the above inequality holds for all

$\gamma > 0$ . We can choose  $\gamma^* = \arg \sup_{\gamma > 0} \gamma(\frac{1}{t}+b) - \Lambda_i(\gamma)$ . Note that  $\gamma^*$  is finite because, according to [10, Lemma 1], if  $\gamma^*$  were infinite, the outage probability of PAoI would be zero, which is not feasible in practice. Substituting  $\gamma^*$  into the inequality, we can obtain:

$$\begin{aligned} \mathbb{P}[A_{i,\infty} \geq x] &\leq e^{2\gamma^* b} \sum_{t \in \{\frac{1}{x}, \frac{2}{x}, \dots\}} e^{-xt(\gamma^*(\frac{1}{t}+b)-\Lambda_i(\gamma^*))} \\ &\stackrel{(a)}{=} e^{2\gamma^* b} \sum_{t \in \{\frac{1}{x}, \frac{2}{x}, \dots\}} e^{-xtI_i(\frac{1}{t}+b)} \\ &\stackrel{(b)}{\leq} e^{2\gamma^* b} \hat{t} x e^{-x \inf_{t>0} tI_i(\frac{1}{t}+b)} \text{ for some constant } \hat{t}, \end{aligned}$$

where step (a) applies the definition of the rate function  $I_i(x)$  and the equality  $\sup_{\gamma \in \mathbf{R}} \gamma x - \Lambda_i(\gamma) = \sup_{\gamma > 0} \gamma x - \Lambda_i(\gamma)$  from [15, Page 303], and step (b) bounds the sum of terms beyond  $\hat{t}x - 1$  (which is a geometric series) by the maximum of the first  $\hat{t}x - 1$  terms, following the bounding technique in [15, Page 315].

Next, we derive a lower bound for the outage probability:

$$\begin{aligned} \mathbb{P}[A_{i,\infty} \geq x] &= \mathbb{P}\left[\max_{s \geq 1} \left\{\sum_{k=1}^s T_{i,k} - sb + 2b\right\} \geq x\right] \\ &\geq \mathbb{P}\left[\sum_{k=1}^s T_{i,k} - sb + 2b \geq x\right] \text{ for any } s \\ &\geq \mathbb{P}\left[\sum_{k=1}^s T_{i,k} - sb \geq x\right] \\ &\stackrel{(a)}{\geq} e^{-x \inf_{t>0} \frac{1}{t} I_i(t+b)}, \end{aligned}$$

where step (a) is from the Cramer–Chernoff theorem as in [15, Page 309].

Combining these bounds, we can express the exponent of the outage probability as

$$\lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} = -\inf_{t>0} \frac{1}{t} I_i(t+b).$$

Finally, we can establish the equivalence as follows:

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{\ln \mathbb{P}[A_{i,\infty} \geq x]}{x} &\leq -\theta_i \\ \iff \inf_{t>0} \frac{1}{t} I_i(t+b) &\geq \theta_i \\ \iff \frac{1}{t} I_i(t+b) &\geq \theta_i \text{ for all } t > 0 \\ \stackrel{(a)}{\iff} t\theta_i - I_i(t+b) &\leq 0 \text{ for all } t > 0 \text{ and also } t \leq 0 \\ \iff \theta_i(t+b) - I_i(t+b) &\leq \theta_i b \text{ for all } t \in \mathbf{R} \\ \iff \max_{t+b \in \mathbf{R}} \theta_i(t+b) - I_i(t+b) &\leq \theta_i b \\ \stackrel{(b)}{\iff} \Lambda_i(\theta_i) &\leq \theta_i b, \end{aligned}$$

where step (a) is because  $I(x) \geq 0$  for all  $x \in \mathbf{R}$ , step (b) uses the property between the rate function  $I_i(x)$  and the LMFG  $\Lambda_i(\gamma)$ , as in [15, Theorem 10.3.4].

## REFERENCES

- [1] X. Qin, Y. Xia, H. Li, Z. Feng, and P. Zhang, "Distributed data collection in age-aware vehicular participatory sensing networks," vol. 8, no. 19, pp. 14 501–14 513, 2021.
- [2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" *Proc. of IEEE INFOCOM*, pp. 2731–2735, 2012.
- [3] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," vol. 65, no. 3, pp. 1807–1827, 2018.
- [4] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," vol. 62, no. 4, pp. 1897–1910, 2016.
- [5] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," vol. 26, no. 6, pp. 2637–2650, 2018.
- [6] T. Park, W. Saad, and B. Zhou, "Centralized and distributed age of information minimization with nonlinear aging functions in the internet of things," vol. 8, no. 10, pp. 8437–8455, 2020.
- [7] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," *Proc. of WiOpt*, pp. 1–8, 2019.
- [8] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," vol. 39, no. 5, pp. 1183–1210, 2021.
- [9] I. Kahraman, A. Köse, M. Koca, and E. Anarim, "Age of information in internet of things: A survey," vol. 11, no. 6, pp. 9896–9914, 2023.
- [10] J.-B. Seo and J. Choi, "On the outage probability of peak age-of-information for d/g/l queuing systems," vol. 23, no. 6, pp. 1021–1024, 2019.
- [11] L. Hu, Z. Chen, Y. Dong, Y. Jia, L. Liang, and M. Wang, "Status update in iot networks: Age-of-information violation probability and optimal update rate," vol. 8, no. 14, pp. 11 329–11 344, 2021.
- [12] A. Zhong, Z. Li, D. Wu, T. Tang, and R. Wang, "Stochastic peak age of information guarantee for cooperative sensing in internet of everything," vol. 10, no. 17, pp. 15 186–15 196, 2023.
- [13] P. D. Mankar, Z. Chen, M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "Throughput and age of information in a cellular-based iot network," vol. 20, no. 12, pp. 8248–8263, 2021.
- [14] R. V. Bhat, R. Vaze, and M. Motani, "Throughput maximization with an average age of information constraint in fading channels," vol. 20, no. 1, pp. 481–494, 2020.
- [15] R. Srikant and L. Ying, *Communication networks: An optimization, control and stochastic networks perspective*. Cambridge University Press, 2014.
- [16] G. De Veciana and J. Walrand, "Effective bandwidths: Call admission, traffic policing and filtering for atm networks," *Queueing Systems*, vol. 20, pp. 37–59, 1995.
- [17] F. B. Hildebrand, *Introduction to numerical analysis*. Courier Corporation, 1987.