A Worst-Case Approximate Analysis of Peak Age-of-Information Via Robust Queueing Approach

Zhongdong Liu, Yu Sang, Bin Li, and Bo Ji

Abstract—A new timeliness metric, called *Age-of-Information (AoI)*, has recently attracted a lot of research interests for real-time applications with information updates. It has been extensively studied for various queueing models based on the probabilistic approaches, where the analyses heavily depend on the properties of specific distributions (e.g., the memoryless property of the exponential distribution or the *i.i.d.* assumption). In this work, we take an alternative new approach, the robust queueing approach, to analyze the *Peak Age-of-Information (PAoI)*. Specifically, we first model the uncertainty in the stochastic arrival and service processes using uncertainty sets. This enables us to approximate the expected PAoI performance for very general arrival and service processes, including those exhibiting heavy-tailed behaviors or correlations, where traditional probabilistic approaches cannot be applied. We then derive a new bound on the PAoI in the single-source single-server setting. Furthermore, we generalize our analysis to two-source single-server systems with symmetric arrivals, which involves new challenges (e.g., the service times of the updates from two sources are coupled in one single uncertainty set). Finally, through numerical experiments, we show that our new bounds provide a good approximation for the expected PAoI. Compared to some well-known bounds in the literature (e.g., one based on Kingman's bound under the *i.i.d.* assumption) that tends to be inaccurate under light load, our new approximation is accurate under both light and high loads, both of which are critical scenarios for the AoI performance.

I. INTRODUCTION

The last decades have witnessed significant advances in the computing capabilities of mobile and sensing devices. The communication capacity of networks has also increased by orders of magnitude. These developments have spawned a wide variety of real-time applications that require timely information updates. A practically important example of such emerging applications is autonomous vehicular systems, where real-time vehicular status information (location, velocity, acceleration, etc.) needs to be shared with nearby vehicles to enable full self-driving capability [1], [2]. Other examples include sensor networks for environmental monitoring, weather or news update applications, and live streaming services.

For such real-time services that require timely information updates, a major concern is about the freshness of the data delivered to the receiver. Those commonly used metrics, such as throughput and delay, cannot precisely measure this timeliness related feature [3]. To that end, in the seminal work [1], a timeliness metric called *Age-of-Information (AoI)* is proposed to measure the freshness of the received data. It is defined as the time elapsed since the most recently received update was generated (see Eq. (2) for a formal definition). In this work, we focus on the metric of *Peak Age-of-Information (PAoI)* [4], which is defined as the maximum value of the AoI before it drops due to a newly delivered fresh update. The PAoI is a critical metric for certain time-sensitive applications that have very stringent timeliness requirements. Consider autonomous vehicular systems, where the status information must be shared with nearby vehicles. The PAoI has to be strictly lower than a certain threshold at all times so that the autonomous vehicular controller makes correct decisions and thus ensures safety and efficiency. Clearly, in such applications, it is insufficient to guarantee a certain level of average AoI only.

Most prior work takes a probabilistic approach to analyze and optimize the AoI and PAoI, which is based on the assumption that the interarrival time and service time follow certain distributions. For example, the exponential distribution has played a privileged role in modeling stochastic systems [3]–[12]. In order to make the analysis tractable, the service times are often assumed to be identically and independently distributed (*i.i.d.*). Although these assumptions lead to tractable performance analysis and optimization, such assumptions may not hold in many practical scenarios. On the other hand, general distributions introduce significant challenges to near-exact analysis of the system performance. It becomes even more challenging if the interarrival times and service times have heavy-tailed distributions and are potentially interdependent.

To that end, we take an alternative approach to model queueing systems based on *robust optimization* [13] and *robust queueing theory* [14], which is originally developed for approximating the system time. Using this new analytical framework, we model the uncertainty in the stochastic arrival and service processes using *uncertainty sets* and approach the problem using a robust optimization formulation. Note that the robust optimization theory can be used to model both light-tailed and heavy-tailed systems. Therefore, our analysis no longer relies on the assumption of specific distributions with attractive properties (e.g., the memoryless property or even *i.i.d.* arrival and service processes). Instead, only the first and second order statistical information (i.e., mean interarrival/service time and variance) is required for the analysis. For the single-source single-server system, we derive an upper bound on the worst-case system time under the assumption

of uncertainty sets, which will be used to approximate the expected PAoI performance. *While Kingman's bound [15], [16] is also a well-known approximation of the system time, it requires i.i.d. interarrival and/or service times. Also, both Kingman's bound and the original robust queueing analysis in [14] are accurate only when the traffic load is high. In contrast, our approach does not have such limitations.* Furthermore, we generalize the analysis to the two-source single-server setting with symmetric traffic arrivals. In this scenario, the updates from two different sources will be processed by a single server in a shared manner, which makes the PAoI analysis more challenging.

We summarize our main contributions as follows.

- *To the best of our knowledge, this is the first work that applies the robust-queueing approach to analyzing the PAoI performance in information-update systems.* This new analytical framework can be applied to a wide range of queueing models without any assumption of specific distributions. In particular, it works well for the systems with non-*i.i.d.* interarrival times and service times.
- We consider a single-source system and derive an upper bound on the PAoI. The upper bound can be used to develop approximations that are very close to the expected PAoI under both light and high traffic loads. This is particularly important to the AoI analysis, as both long service times (when the load is high) and long interarrival times (when the load is light) would result in a large AoI.
- We further generalize the analysis to the two-source setting with symmetric arrivals. The generalization is non-trivial and involves new challenges. One key challenge is that the service times of the updates from two sources are coupled in one single uncertainty set. Therefore, the property of uncertainty set cannot be directly used for analyzing the PAoI performance of each source.
- Finally, we perform extensive numerical experiments and evaluate the PAoI performance under different traffic loads as well as for different stochastic processes. The simulation results show that our new bound with properly chosen parameters of uncertainty sets provides accurate approximations for the PAoI performance.

The remainder of this paper is organized as follows. We discuss the related work on AoI and robust queueing theory in Section II. Then, we describe our model and provide our analysis for a single-source system in Section III. In Section IV, we generalize our results to the two-source case. Finally, we present the numerical results in Section V and make concluding remarks in Section VI.

II. RELATED WORK

The AoI, a recently proposed metric, has inspired a series of studies on the analysis and optimization of the timeliness performance (see [17]–[19] for a survey). The notion of AoI is formally introduced in [3], where the authors analyze the time-average AoI in M/M/1, M/D/1, and D/M/1 systems under the First-Come-First-Served (FCFS) policy. In [5], the average AoI

is analyzed for the M/M/1 system under the Last-Come-First-Served (LCFS) policy with and without preemption. In [4], the AoI performance of the FCFS policy in the M/M/1/1 and M/M/1/2 queues is studied, where new arrivals are discarded if the buffer is full. More sophisticated models have also been considered in the literature, such as two-source systems [8]–[10] and multi-server systems [7], [20], [21]. However, most of the previous studies adopt the traditional probabilistic analytical framework and assume that the interarrival time and service time follow certain distributions.

In order to overcome the limitations of the probabilistic framework, recent work also considers other approaches to calculate the average AoI. For example, the authors in [22] derive the stationary distribution of the AoI, which is in terms of the stationary distribution of the delay and the PAoI. With the AoI distribution, one can analyze the mean or higher moments of the AoI in GI/GI/1, M/GI/1, and GI/M/1 queues under several scheduling policies (e.g., FCFS and LCFS). The authors in [23] characterize the violation probability of AoI and use it to obtain upper bounds of AoI for GI/GI/1/1 and GI/GI/1/2* systems. However, these studies are still based on the assumption of *i.i.d.* random variables, which is not required in our approach. In [24], the stochastic hybrid system (SHS) is introduced as an analytical technique for studying the AoI. The SHS method provides a way to derive closed-form AoI results for simple queues described as finite-state Markov chains. However, the SHS-based technique is limited to finite-state systems only (e.g., finite-buffer queueing systems) and is inapplicable to our problem.

In this paper, we propose an alternative approach based on the robust queueing framework to analyze the AoI. In what follows, we briefly discuss recent developments of robust optimization and robust queueing theory. Robust optimization has been proven to be an efficient approach for complex optimization problems with significant uncertainties [13], [25]. This approach is later adopted in the development of robust queueing theory [14], which can be used to provide fairly accurate predictions of the performance of complex queueing systems without making any probabilistic assumptions. Note that in the traditional probabilistic approaches, it is often assumed that the underlying queueing system has the Markovian property (i.e., exponentially distributed interarrival times and service times) and that the interarrival times and the service times are both *i.i.d.* In contrast, using the robust-queueing approach, we no longer make such probabilistic assumptions. Motivated by the Central Limit Theorem (CLT) and its generalized version, the randomness is modeled as uncertainty sets rather than specific probabilistic distributions. Moreover, the interarrival times and service times do not have to be *i.i.d.* over time [14]. This enables us to characterize the queueing performance for very general arrival and service processes, including those exhibiting heavy-tailed behaviors or correlations. In this paper, we will adopt this framework for the PAoI analysis in information-update systems.

Fig. 1: A single-source information-update system.

III. SINGLE-SOURCE SYSTEM

In this section, we discuss a simple case with one source, one server and one monitor. We then generalize the analysis to the setting with two pairs of source and monitor in Section IV.

A. System Model

In a single-source information-update system, there is only one pair of source and monitor as illustrated in Fig. 1. Each update is stamped with the time when it is generated. Let a_n denote the generation time of the n-th update. Then the interarrival time between the n-th update and the $(n-1)$-st update can be denoted by $T_n \triangleq a_n - a_{n-1}$. We assume that the interarrival time follows a general distribution with mean $1/\lambda$. After an update is generated, it needs to be processed by the server before it is delivered to the monitor. The server has a FCFS queue of infinite buffer size. We assume that the update arrives to the queue immediately after being generated. Hence, the n-th update arrives to the queue at time a_n as well. Let W_n denote the waiting time of the n-th update. The service time of the n-th update is denoted by X_n, which also follows a general distribution with mean $1/\mu$. Let f_n denote the time when the service of the n-th update at the server is finished. After the update completes its service, it will be immediately delivered to the monitor. Therefore, the update arrives at the monitor also at time f_n. Let S_n denote the total system time experienced by the n-th update, which is also equal to the sum of its waiting time in the queue and the service time,

$$S_n \triangleq f_n - a_n = W_n + X_n. \tag{1}$$

For such a time-sensitive information-update system, we are interested in the freshness of data at the monitor. The freshness of data is measured by the metric *Age-of-Information* (AoI). It is defined as the time elapsed since the freshest update received by the monitor was generated. Assume that the latest update received by the monitor at time t is stamped with the generation time $a(t)$, then the AoI at the monitor is

$$\Delta(t) \triangleq t - a(t). \tag{2}$$

An example of the evolution of the AoI at the monitor is shown in Fig. 2. The AoI would increase linearly when no update is transmitted to the monitor, and it reaches a local maximum value immediately before a new update is delivered. Such a maximum value is called the *Peak Age-of-Information (PAoI)*. There is a PAoI corresponding to each update in our example. Let P_n denote the n-th PAoI. From Fig. 2, it is easy to see $P_n = f_n - a_{n-1}$. This can be rewritten as the sum of the interarrival time between the n-th and $(n-1)$-st updates (i.e., $T_n = a_n - a_{n-1}$) and the system time of the n-th update (i.e., $S_n = f_n - a_n$). Then, the expected PAoI can be expressed as

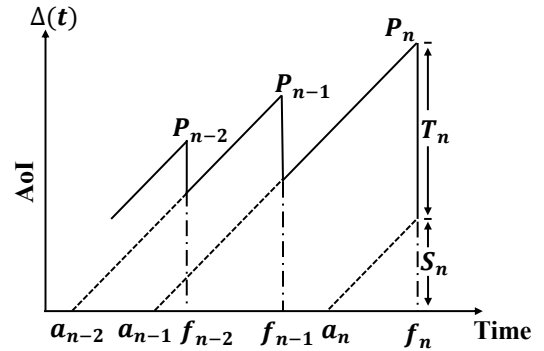$$\mathbb{E}[P_n] = \mathbb{E}[T_n] + \mathbb{E}[S_n]. \tag{3}$$

Fig. 2: An illustration of the AoI evolution at the monitor in the single-source system.

It is important to study the PAoI performance since it represents the locally largest values of the AoI at the monitor and captures more stringent timeliness requirements. Most of the previous work takes a probabilistic approach to analyze the PAoI, which assumes specific types of distribution of the service and interarrival times. Taking a different path, we adopt a worst-case approximation approach based on the robust queueing theory, which only requires the knowledge of the first and second order statistical information of the distributions.

B. The Worst-Case Approach

In this subsection, we will analyze the expected PAoI (see Eq. (3)). Since we trivially have $\mathbb{E}[T_n] = 1/\lambda$, it remains to analyze the expected system time $\mathbb{E}[S_n]$.

Similar to [14], we consider a sample path of n updates. The generation and service process of the updates can be characterized by the interarrival times $\mathbf{T}_n \triangleq (T_1, T_2, \ldots, T_n)$ and the service times $\mathbf{X}_n \triangleq (X_1, X_2, \ldots, X_n)$, respectively. Applying the robust queueing theory proposed in [14], we adopt a worst-case analysis to obtain an upper bound on the system time S_n, which will be leveraged to approximate the expected system $\mathbb{E}[S_n]$. *Different from [14] that neglects the accuracy of S_n under light load, we characterize S_n under both light and high loads in the single-source system and the two-source system (see more discussions in Section III-C and Section IV, respectively).* To begin with, we express the system time of the n-th update S_n as follows [26]:

$$S_n = W_n + X_n$$
$$= \max_{1 \le k \le n} \left(\sum_{i=k}^{n} X_i - \sum_{i=k+1}^{n} T_i \right), \tag{4}$$

where we recall that W_n (X_n, resp.) is the waiting time (the service time, resp.) of the n-th update. Note that S_n involves interarrival times $\mathbf{T}_n$ and service times $\mathbf{X}_n$.

From Eq. (4), we can see that the analysis of the system time involves the sum of system time and interarrival time of previous updates, which often makes the probabilistic approach intractable. On the other hand, the Generalized Central Limit Theorem (GCLT) indicates that the distribution of the sum of multiple *i.i.d.* random variables converges to a stable

distribution. Motivated by GCLT, we assume that the partial sum of service times satisfies the following:

$$\frac{\sum_{i=k}^{n} X_i - (n-k+1)/\mu}{(n-k+1)^{1/\alpha}} \leq \Gamma_s, \tag{5}$$

where $1/\mu$ is the expected service time, $\alpha \in (1,2]$ is the tail coefficient that models possibly heavy-tailed probability distributions[1] (the closer to 1, the heavier the tail), and $\Gamma_s > 0$ is the variability parameter chosen to ensure that Eq. (5) is satisfied with high probability. For instance, the normalized sum of a large number of positive Pareto random variables can be approximated by a random variable Y following a standard stable distribution with a tail coefficient α and $C_\alpha = [G(1-\alpha)\cos(\pi\alpha/2)]^{1/\alpha}$, where $G(\cdot)$ represents the gamma function. For a tail coefficient of $\alpha = 1.5$, we obtain $\mathbb{P}(Y \leq 6.5) \approx 0.975$ and $\mathbb{P}(Y \leq 19) \approx 0.995$ [14]. Thus, choosing $\Gamma_s = 6.5$ is one good option in this instance. Note that there is a tradeoff between robustness and accuracy when choosing Γ_s. That is, when choosing a larger enough Γ_s, we can ensure that Eq. (5) is satisfied with a higher probability (i.e., stronger robustness), but this may result in a loose bound on the system time (i.e., lower accuracy). On the other hand, when choosing a smaller Γ_s, we can obtain tight bounds on the system time, but Eq. (5) may be violated with a higher probability. In Section V-A, we discuss one way of choosing Γ_s that attempts to balance the tradeoff. With a properly chosen Γ_s, we further assume that the service times $\mathbf{X}_n$ belong to the following parameterized uncertainty set [14]:

$$\mathcal{U}_s \triangleq \left\{ \mathbf{X}_n \mid \frac{\sum_{i=k}^{n} X_i - (n-k+1)/\mu}{(n-k+1)^{1/\alpha}} \leq \Gamma_s, \right.$$
$$\left. \forall\, 1 \leq k \leq n \right\}. \tag{6}$$

Although the uncertainty set is motivated by the *i.i.d.* assumption, $\mathbf{X}_n \in \mathcal{U}_s$ does not necessarily require that $X_1, X_2, \ldots, X_n$ be *i.i.d.*

Similarly, we assume that the interarrival times $\mathbf{T}_n$ belong to the following uncertainty set:

$$\mathcal{U}_a \triangleq \left\{ \mathbf{T}_n \mid \frac{\sum_{i=k+1}^{n} T_i - (n-k)/\lambda}{(n-k)^{1/\alpha}} \geq -\Gamma_a, \right.$$
$$\left. \forall\, 0 \leq k \leq n-1 \right\}, \tag{7}$$

where $1/\lambda$ is the mean interarrival time, $\alpha \in (1,2]$ is the tail coefficient that models possibly heavy-tailed probability distributions, and $\Gamma_a > 0$ is the variability parameter chosen to ensure that Eq. (7) is satisfied with high probability. We consider the lower bound on the partial sums of the interarrival times as that leads to the worst-case system time. We assume that the distributions of the service time and interarrival time have the same tail coefficient α. However, we assume no dependence between $\mathbf{T}_n$ and $\mathbf{X}_n$ and will discuss the case with dependence towards the end of this paper.

[1]The heavy-tailed distributions (like the Pareto or the Weibull) have heavier tails than the exponential distribution. Roughly speaking, there is a larger probability of getting very large values.

C. Performance Analysis

Let $\hat{S}_n$ denote the worst-case system time of the n-th update. From Eq. (4), we have the following:

$$\hat{S}_n \triangleq \max_{\mathbf{T}_n \in \mathcal{U}_a} \max_{\mathbf{X}_n \in \mathcal{U}_s} \max_{1 \leq k \leq n} \left(\sum_{i=k}^{n} X_i - \sum_{i=k+1}^{n} T_i \right)$$
$$\leq \max_{1 \leq k \leq n} \left(\max_{\mathbf{X}_n \in \mathcal{U}_s} \sum_{i=k}^{n} X_i - \min_{\mathbf{T}_n \in \mathcal{U}_a} \sum_{i=k+1}^{n} T_i \right). \tag{8}$$

It is shown in [14] that we can find the sample path of $\hat{\mathbf{X}}_n \in \mathcal{U}_s$ and $\hat{\mathbf{T}}_n \in \mathcal{U}_a$ that achieve the worst-case system time (i.e., $\sum_{i=k}^{n} \hat{X}_i = \max_{\mathbf{X}_n \in \mathcal{U}_s} \sum_{i=k}^{n} X_i$ and $\sum_{i=k+1}^{n} \hat{T}_i = \min_{\mathbf{T}_n \in \mathcal{U}_a} \sum_{i=k+1}^{n} T_i$). Plugging such $\hat{\mathbf{T}}_n$ and $\hat{\mathbf{X}}_n$ into Eq. (8) gives the exact expression of $\hat{S}_n$:

$$\hat{S}_n = \max_{1 \leq k \leq n} \left\{ \frac{n-k+1}{\mu} - \frac{n-k}{\lambda} + \Gamma_s(n-k+1)^{1/\alpha} \right.$$
$$\left. + \Gamma_a(n-k)^{1/\alpha} \right\}. \tag{9}$$

As we can see from Eq. (9), the worst-case system time $\hat{S}_n$ is proportional to the mean service time $1/\mu$ and inversely proportional to the mean interarrival time $1/\lambda$. Also, larger Γ_s and Γ_a lead to a larger $\hat{S}_n$, given that Γ_s and Γ_a are used to bound the sum of service times $\mathbf{X}_n$ and the sum of interarrvial times $\mathbf{T}_n$ in the uncertainty sets, respectively.

Based on Eq. (9) and some relaxations, one can derive an upper bound on the worst-case system time $\hat{S}_n$. We restate this result in the following lemma.

Lemma 1 (Theorem 2 in [14]). *In a single-source FCFS queueing system with* $\mathbf{T}_n \in \mathcal{U}_a$ *and* $\mathbf{X}_n \in \mathcal{U}_s$, *we have*

$$\hat{S}_n \leq \frac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \frac{(\Gamma_s + \Gamma_a)^{\alpha/(\alpha-1)}}{(1/\lambda - 1/\mu)^{1/(\alpha-1)}} + 1/\lambda. \tag{10}$$

It has been shown that the above upper bound is nearly tight when the traffic load is high [14]. However, the relaxations used in [14] renders the bound loose when the traffic load is light, i.e., when λ is relatively small. This is acceptable in the analysis of queueing system with respect to traditional metrics, such as delay or throughput, which are typically pronounced in the high-load regime. *While for the metric of PAoI, a low arrival rate means long interarrival times between consecutive updates, which leads to a poor PAoI performance as well. This indicates that both light and high loads affect the PAoI performance significantly, and thus, focusing on the system time in the high-load regime only is insufficient.* Instead of introducing relaxations, we propose an alternative method that provides the exact characterization of the worst-case system time. We derive a new upper bound of the worst-case system time, which is nearly tight under both light and high loads. The upper bound is presented in the following theorem.

Theorem 1. *Define the following function* $f(m) \triangleq (m+1)/\mu - m/\lambda + \Gamma_s(m+1)^{1/\alpha} + \Gamma_a m^{1/\alpha}$ *and let*

$$l = \left(\frac{\alpha(1/\lambda - 1/\mu)}{\Gamma_a + \Gamma_s} \right)^{\alpha/(1-\alpha)}, \tag{11}$$

where $\alpha \in (1, 2]$. *In a single-server FCFS queueing system with $\mathbf{T}_n \in \mathcal{U}_a$ and $\mathbf{X}_n \in \mathcal{U}_s$, we have*

$$\hat{S}_n \leq \begin{cases} \max\{f(n-1), 0\}, & \text{if } n - 1 \leq \lfloor l \rfloor - 1 \\ \max\{f(m^*), 0\}, & \text{otherwise} \end{cases}, \quad (12)$$

where $m^* \in \underset{m \in \{\lfloor l \rfloor - 1, \lfloor l \rfloor, \lfloor l \rfloor + 1\} \cap [0, n-1]}{\arg \max} f(m).$

Proof. The derivation of the upper bound of the worst-case system time $\hat{S}_n$ in Eq. (12) utilizes the convexity of Eq. (9). Let $m = n - k$. According to Eq. (9), we can rewrite the worst-case system time as

$$\hat{S}_n = \max_{0 \leq m \leq n-1} \frac{m+1}{\mu} - \frac{m}{\lambda} + \Gamma_s(m+1)^{1/\alpha} + \Gamma_a m^{1/\alpha}$$

$$= \max_{0 \leq m \leq n-1} f(m). \quad (13)$$

Let m^* be the integral maximizer of $f(m)$, i.e., $\hat{S}_n = f(m^*)$. In order to find the bound of m^*, we extend the domain of $f(m)$ to the set of nonnegative real numbers (i.e., $m \in R^+$). The second order derivative of $f(m)$ is

$$f''(m) = \frac{1 - \alpha}{\alpha^2} \Big(\Gamma_s(m+1)^{(1-2\alpha)/\alpha} + \Gamma_a m^{(1-2\alpha)/\alpha} \Big), \quad (14)$$

which is negative since $\alpha \in (1, 2]$. This implies that $f(m)$ is concave. Let $M \in R^+$ be the continuous maximizer of $f(m)$, i.e., M is the solution of the following equation:

$$f'(m) = \frac{1}{\alpha} \Big(\Gamma_s(m+1)^{\frac{1-\alpha}{\alpha}} + \Gamma_a m^{\frac{1-\alpha}{\alpha}} \Big) - \frac{1}{\lambda} + \frac{1}{\mu} = 0. \quad (15)$$

However, it is usually difficult to solve Eq. (15) to get the expression of M. We then define function $g(m)$ as

$$g(m) = \frac{1}{\alpha} (\Gamma_s m^{\frac{1-\alpha}{\alpha}} + \Gamma_a m^{\frac{1-\alpha}{\alpha}}) - \frac{1}{\lambda} + \frac{1}{\mu}, \quad (16)$$

and let l be the solution of $g(m) = 0$, which gives

$$g(l) = \frac{1}{\alpha} (\Gamma_s l^{\frac{1-\alpha}{\alpha}} + \Gamma_a l^{\frac{1-\alpha}{\alpha}}) - \frac{1}{\lambda} + \frac{1}{\mu} = 0, \quad (17)$$

and

$$l = \left(\frac{\alpha(1/\lambda - 1/\mu)}{\Gamma_a + \Gamma_s} \right)^{\frac{\alpha}{1-\alpha}}. \quad (18)$$

Note that the following is satisfied:

$$f'(l) = \frac{1}{\alpha} \Big(\Gamma_s(l+1)^{\frac{1-\alpha}{\alpha}} + \Gamma_a l^{\frac{1-\alpha}{\alpha}} \Big) - \frac{1}{\lambda} + \frac{1}{\mu}$$

$$= \frac{\Gamma_s}{\alpha} \Big((l+1)^{\frac{1-\alpha}{\alpha}} - l^{\frac{1-\alpha}{\alpha}} \Big)$$

$$< 0, \quad (19)$$

where the second equality follows from Eq. (17) and the last inequality holds because $x^{\frac{1-\alpha}{\alpha}}$ is a decreasing function for $\alpha \in (1, 2]$. Similarly, we can show $f'(l-1) > 0$. Therefore, we have $f'(l-1) > f'(M) > f'(l)$, which implies $l-1 < M < l$ since $f'(m)$ is a decreasing function due to the concavity of $f(m)$. Some thoughts give that the integral maximizer m^* must satisfy the following: $m^* \in \{\lfloor l \rfloor - 1, \lfloor l \rfloor, \lfloor l \rfloor + 1\}$. Recall that function $f(m)$ is defined on $[0, n-1]$. If $n - 1 \leq \lfloor l \rfloor - 1$, i.e., the integral maximizer m^* is out of the domain, then we have $\hat{S}_n \leq f(n-1)$ since $f(m)$ is an increasing function on $[0, n-1]$; otherwise, if $n - 1 > \lfloor l \rfloor - 1$, then m^* must be a value belonging to $\{\lfloor l \rfloor - 1, \lfloor l \rfloor, \lfloor l \rfloor + 1\}$ that satisfies
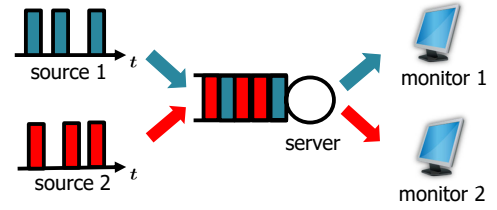

Fig. 3: A two-source information-update system.

$m^* \leq n - 1$ and achieves the maximum of $f(m)$. Finally, we take the maximum between $f(m^*)$ and 0 because the system time is nonnegative. This completes the proof. $\square$

Using the upper bound of the worst-case system time derived in Theorem 1, we can approximate the expected system time by choosing appropriate variability parameters Γ_a and Γ_s (see Section V-A for detailed discussions). This enables us to approximate the expected PAoI given that the expected interarrival time is already known (i.e., $1/\lambda$).

IV. TWO-SOURCE SYSTEM

In this section, we consider a more general system that consists of two sources, one shared server, and two monitors. Each monitor receives the update only from one source. An illustration of the system is shown in Fig. 3. Our interest is to analyze the expected PAoI performance at each monitor. Note that this generalization is non-trivial and involves new technical challenges (e.g., the service times of the updates from two sources are coupled in one single uncertainty set).

A. System Model

There are a total number of n updates from both sources all going through one single server to the corresponding monitor. We assume that these two sources are symmetric, i.e., their update interarrival time follows the same general distribution with mean $1/\lambda$ and their update's service time also follows the same general distribution with mean $1/\mu$. We assume $2\lambda/\mu < 1$ in order to keep the system stable. We use $a_n^{(s)}$ ($f_n^{(s)}$, resp.) to denote the arrival time (service completion time, resp.) of the n-th update from source $s = 1, 2$. The service time of the n-th update from source s is denoted by $X_n^{(s)} \triangleq f_n^{(s)} - a_n^{(s)}$. The interarrival time between the n-th update and the $(n-1)$-st update from source s is denoted by $T_n^{(s)} \triangleq a_n^{(s)} - a_{n-1}^{(s)}$. Then, the AoI at monitor s is defined as

$$\Delta^{(s)}(t) \triangleq t - a^{(s)}(t), \quad (20)$$

where $a^{(s)}(t)$ denotes the generation time of the most recently received update at monitor s. Similarly, the n-th PAoI at monitor s can be expressed as $P_n^{(s)} \triangleq f_n^{(s)} - a_{n-1}^{(s)} = T_n^{(s)} + S_n^{(s)}$, and the expected PAoI can be denoted as

$$\mathbb{E}[P_n^{(s)}] = \mathbb{E}[T_n^{(s)}] + \mathbb{E}[S_n^{(s)}] = 1/\lambda + \mathbb{E}[S_n^{(s)}]. \quad (21)$$

Similar to the single-source case, we utilize the robust queueing theory to analyze the worst-case performance of S_n (the system time of the n-th update, either from source 1 or 2) and use it to approximate $\mathbb{E}[S_n^{(s)}]$.

Since the updates from both sources will join the same FCFS queue, we reorder all updates according to their arrival time. Let a_n denote the arrival time of the n-th update that arrives at the server[2]. Note that this update could be from either source 1 or 2. Consider a sample path of n updates that arrive at the server at time $\mathbf{A}_n = \{a_1, a_2, \ldots, a_n\}$ such that $a_1 \le a_2 \le \cdots \le a_n$. The service time of the n-th update is denoted by X_n, and corresponding service times for the whole sample path are $\mathbf{X_n} = \{X_1, X_2, \ldots, X_n\}$. The interarrival time between the n-th and $(n-1)$-st updates is denoted by $T_n \triangleq a_n - a_{n-1}$, and the corresponding interarrival times for the whole sample path are $\mathbf{T}_n = \{T_1, T_2, \ldots, T_n\}$. To distinguish the updates from source 1 and source 2, we further define the following functions that map the update to its source according to the arrival times:

$$b_s(n) \triangleq \arg\min_k a_k^{(s)} \ge a_n, \ e_s(n) \triangleq \arg\max_k a_k^{(s)} \le a_n,$$
$$s = 1, 2.$$

According to the mapping function, $e_s(n)$ is the index of the last update from source s and $a_{e_s(n)}^{(s)}$ is the arrival time of this last update. Therefore, the interarrival times of two sources can be denoted as $\mathbf{T}_{e_1(n)}^{(1)} = \{T_1^{(1)}, T_2^{(1)}, \ldots, T_{e_1(n)}^{(1)}\}$ and $\mathbf{T}_{e_2(n)}^{(2)} = \{T_1^{(2)}, T_2^{(2)}, \ldots, T_{e_2(n)}^{(2)}\}$, respectively.

Now, we consider the system time corresponding to the last update that arrives at the server (i.e., the n-th update). Assume that this update is from source 1. Recall that the system time S_n (see Eq. (4)) can be written as

$$S_n = \max_{1 \le k \le n} \left(\sum_{i=k}^n X_i - \sum_{i=k+1}^n T_i \right). \quad (22)$$

In the rest of the paper, the analysis is based on the assumption that the n-th update is from source 1. It is easy to apply the same analysis to the case that the n-th update is from source 2.

For the two-source system, the AoI is determined by service times $\mathbf{X}_n$ and the interarrival times of two sources, $\mathbf{T}_{e_1(n)}^{(1)}$ and $\mathbf{T}_{e_2(n)}^{(2)}$. We need three uncertainty sets for them. First, for the service times, we assume that two sources have the same type of updates. Therefore, the same assumption as in the single-source system still holds since we also have only one server, i.e.,

$$\mathcal{U}_s = \left\{ (X_1, X_2, \ldots, X_n) \Big| \frac{\sum_{i=k}^n X_i - \frac{n-k+1}{\mu}}{(n-k+1)^{1/\alpha}} \le \Gamma_s, \right.$$
$$\left. \forall\, 1 \le k \le n \right\}. \quad (23)$$

For the interarrival time $\mathbf{T}_{e_1(n)}^{(1)}$ and $\mathbf{T}_{e_2(n)}^{(2)}$, we assume that the uncertainty set for source s is

$$\mathcal{U}_a^s = \left\{ (T_1^{(s)}, T_2^{(s)}, \ldots, T_{e_s(n)}^{(s)}) \Big| \frac{\sum_{i=k}^l T_i^{(s)} - \frac{l-k+1}{\lambda}}{(l-k+1)^{1/\alpha}} \ge -\Gamma_a, \right.$$
$$\left. \forall\, 1 \le l \le e_s(n) \text{ and } 1 \le k \le l \right\}. \quad (24)$$

Here the assumptions for α, Γ_s, and Γ_a are the same as that in the single-source case. With the uncertainty sets for each source, we further define the uncertainty set for the whole sample path, $\mathcal{U}_a = \mathcal{U}_a^1 \cup \mathcal{U}_a^2$.

B. Performance Analysis

In this subsection, we derive an upper bound on the worst-case system time in the two-source system.

Theorem 2. Define the following function[3] $f(m) \triangleq 2(m+1)/\mu - m/\lambda + 2\Gamma_s(m+1)^{1/\alpha} + \Gamma_a m^{1/\alpha}$, and let

$$l = \left(\frac{\alpha(1/\lambda - 2/\mu)}{\Gamma_a + 2\Gamma_s} \right)^{\frac{\alpha}{1-\alpha}}, \quad (25)$$

where $\alpha \in (1, 2]$. In a single-server FCFS queue with two symmetric sources, such that $\mathbf{T}_{e_s(n)}^{(s)} \in \mathcal{U}_a^s$ ($s = 1, 2$) and $\mathbf{X}_n \in \mathcal{U}_s$, we have

$$\hat{S}_n \le \begin{cases} \max\left\{ f\left(-\frac{1}{2}\right), f\left(\frac{n}{2} - 1\right), 0 \right\}, & \text{if } \frac{n}{2} - 1 \le \lfloor l \rfloor - 1 \\ \max\left\{ f\left(-\frac{1}{2}\right), f(m^*), 0 \right\}, & \text{otherwise} \end{cases},$$
$$(26)$$

where $m^* \in \arg\max\limits_{m \in \mathcal{D} \cap \{m \le n/2 - 1\}} f(m)$ and $\mathcal{D} \triangleq \{\lfloor l \rfloor - 1, \lfloor l \rfloor - 1/2, \lfloor l \rfloor, \lfloor l \rfloor + 1/2, \lfloor l \rfloor + 1\}$.

Proof. As shown in Eq. (26), the expression of the system time is very similar to that in the single-source system. However, the analysis is quite different since the interarrival times consist of two sequences. Given $\mathbf{T}_{e_1(n)}^{(1)}$ and $\mathbf{T}_{e_2(n)}^{(2)}$, the worst-case system time is

$$\hat{S}_n(\mathbf{T}_{e_1(n)}^{(1)}, \mathbf{T}_{e_2(n)}^{(2)}) \le \max_{1 \le k \le n} \left(\max_{\mathbf{X}_n \in \mathcal{U}_s} \sum_{i=k}^n X_i - \sum_{i=k+1}^n T_i \right). \quad (27)$$

First, it is shown in [14] that there exists a sequence of service times $\hat{\mathbf{X}}_n \in \mathcal{U}_s$ that achieves the upper bound in Eq. (27). We restate the results here that the sequence $\hat{\mathbf{X}}_n$ satisfies:

$$\sum_{i=k}^n \hat{X}_i = \max_{\mathbf{X}_n \in \mathcal{U}_s} \sum_{i=k}^n X_i = \frac{n-k+1}{\mu} + \Gamma_s(n-k+1)^{1/\alpha},$$
$$\forall\, k = 1, 2, \ldots, n. \quad (28)$$

This implies that the service times that achieve the worst-case system time are independent of the interarrival times. Therefore, we can replace the partial sum of service times with Eq. (28):

$$\hat{S}_n = \max_{\mathbf{T}_n \in \mathcal{U}_a} \max_{1 \le k \le n} \left(\sum_{i=k}^n \hat{X}_i - \sum_{i=k+1}^n T_i \right)$$
$$\le \max_{1 \le k \le n} \left(\sum_{i=k}^n \hat{X}_i - \min_{\mathbf{T}_n \in \mathcal{U}_a} \sum_{i=k+1}^n T_i \right) \quad (29)$$

Different from the service times, the interarrival times are from two different uncertainty sets and the number of updates from each source depends on each other. There do not exist two interarrival time sequences that achieve the maximum value of the system time for all possible k's as we show in the

single-source case. As such, we define function $S(k)$ for all $k = 1, 2, \ldots, n$ as

$$S(k) \triangleq \sum_{i=k}^{n} \hat{X}_i - \min_{\mathbf{T}_n \in \mathcal{U}_a} \sum_{i=k+1}^{n} T_i, \tag{30}$$

and maximize the value of $S(k)$ for every $1 \leq k \leq n$. Therefore, the worst-case system time can be rewriten as

$$\hat{S}_n = \max_{1 \leq k \leq n} S(k). \tag{31}$$

Note that $S(k)$ is completely determined by the term of interarrival times' partial sum $T(k)$, which is defined as

$$T(k) \triangleq \min_{\mathbf{T}_n \in \mathcal{U}_a} \sum_{i=k+1}^{n} T_i. \tag{32}$$

Then, deriving an upper bound of $S(k)$ is equivalent to deriving a lower bound of $T(k)$. The partial sum of the interarrival times from source 1 and source 2 are both less than $T(k)$, i.e.,

$$T(k) \geq \min_{\mathbf{T}_{e_1(n)}^{(1)} \in \mathcal{U}_a^1} \sum_{i=b_1(k)+1}^{e_1(n)} T_i^{(1)} \text{ and} \tag{33}$$

$$T(k) \geq \min_{\mathbf{T}_{e_2(n)}^{(2)} \in \mathcal{U}_a^2} \sum_{i=b_2(k)+1}^{e_2(n)} T_i^{(2)}. \tag{34}$$

As shown in uncertainty sets $\mathcal{U}_a^1$ and $\mathcal{U}_a^2$, the lower bound of the partial sum of interarrival times depends on the number of updates. The total number of updates from the k-th update to the n-th update is $(n-k+1)$. Let h be the number of updates from source 2, i.e., $e_2(n) - b_2(k) + 1 = h$. Correspondingly, we have $e_1(n) - b_1(k) + 1 = n - k + 1 - h$ updates from source 1. According to the uncertainty sets Eq. (24), we have the following two lower bounds:

$$\sum_{i=b_1(k)+1}^{e_1(n)} T_i^{(1)} \geq \frac{e_1(n) - b_1(k)}{\lambda} - \Gamma_a \left(e_1(n) - b_1(k) \right)^{1/\alpha}$$

$$= \frac{n-k-h}{\lambda} - \Gamma_a (n-k-h)^{1/\alpha}, \tag{35}$$

$$\sum_{i=b_2(k)+1}^{e_2(n)} T_i^{(2)} \geq \frac{e_2(n) - b_2(k)}{\lambda} - \Gamma_a \left(e_2(n) - b_2(k) \right)^{1/\alpha}$$

$$= \frac{h-1}{\lambda} - \Gamma_a (h-1)^{1/\alpha}. \tag{36}$$

We denote the lower bounds in Eq. (35) and Eq. (36) as $L_1(h) = (n-k-h)/\lambda - \Gamma_a(n-k-h)^{1/\alpha}$ and $L_2(h) = (h-1)/\lambda - \Gamma_a(h-1)^{1/\alpha}$, respectively. Therefore, we have

$$T(k) \geq \min_{0 \leq h \leq n-k} \max\{L_1(h), L_2(h)\}. \tag{37}$$

Note that both $L_1(h)$ and $L_2(h)$ are convex functions and they are symmetric along the line of $h = (n-k+1)/2$, which implies that the continuous minimizer of Eq. (37) is $h' = (n-k+1)/2$. Therefore, the lower bound for $T(k)$ is

$$T(k) \geq (n-k-1)/(2\lambda) - \Gamma_a((n-k-1)/2)^{1/\alpha}. \tag{38}$$

By replacing the partial sum $T(k)$ with its lower bound in Eq. (38), we define the upper bound of $S(k)$ as $\hat{S}(k)$, i.e.,

$$\hat{S}(k) = \sum_{i=k}^{n} \hat{X}_i - \frac{n-k-1}{2\lambda} + \Gamma_a \left(\frac{n-k-1}{2\lambda} \right)^{1/\alpha}$$

$$= \frac{n-k+1}{\mu} + \Gamma_s (n-k+1)^{1/\alpha}$$

$$- \frac{n-k-1}{2\lambda} + \Gamma_a \left(\frac{n-k-1}{2} \right)^{1/\alpha}. \tag{39}$$

Therefore, we have $S(k) \leq \hat{S}(k)$ for all $1 \leq k \leq n$. Then, we can rewrite the upper bound of the system time as

$$\hat{S}_n \leq \max_{1 \leq k \leq n} \hat{S}(k). \tag{40}$$

Similar to the proof of Theorem 1, let $m = (n - k - 1)/2 \in \{-1/2, 0, 1/2, \ldots, n/2 - 1\}$. Then, according to Eq.(39) and the defined $f(m)$, Eq.(40) can be rewritten as

$$\hat{S}_n \leq \max_{-\frac{1}{2} \leq m \leq \frac{n}{2} - 1} f(m) = \max \left\{ f\left(-\frac{1}{2} \right), \max_{0 \leq m \leq \frac{n}{2} - 1} f(m) \right\}. \tag{41}$$

Next, we consider the second item in max function of Eq. (41). Let m^* be the maximizer, i.e., $f(m^*) = \max_{0 \leq m \leq \frac{n}{2} - 1} f(m)$. We can also extend $m \in \{0, 1/2, \ldots, n/2 - 1\}$ to the real numbers, i.e., $m \in R^+$. It is easy to check that $f(m)$ is also concave since we have

$$f''(m) = \frac{1-\alpha}{\alpha^2} \left(2\Gamma_s(m+1)^{\frac{1-2\alpha}{\alpha}} + \Gamma_a m^{\frac{1-2\alpha}{\alpha}} \right) < 0. \tag{42}$$

Let $M \in R^+$ be the continuous maximizer of $f(m)$, i.e.,

$$f'(M) = \frac{2}{\mu} - \frac{1}{\lambda} + \frac{1}{2} \left(2\Gamma_s(M+1)^{\frac{1-\alpha}{\alpha}} + \Gamma_a M^{\frac{1-\alpha}{\alpha}} \right) = 0. \tag{43}$$

In order to obtain the range of M, let l be the solution of the following equation,

$$\frac{2}{\mu} - \frac{1}{\lambda} + \frac{1}{\alpha} \left(2\Gamma_s m^{\frac{1-\alpha}{\alpha}} + \Gamma_a m^{\frac{1-\alpha}{\alpha}} \right) = 0, \tag{44}$$

which gives

$$l = \left(\frac{\alpha(1/\lambda - 2/\mu)}{\Gamma_a + 2\Gamma_s} \right)^{\frac{\alpha}{1-\alpha}}. \tag{45}$$

Similar to the single-source case, we have $l - 1 < M < l$. Therefore, there must be $m^* \in \{\lfloor l \rfloor - 1, \lfloor l \rfloor - 1/2, \lfloor l \rfloor, \lfloor l \rfloor + 1/2, \lfloor l \rfloor + 1\}$. Recall that the concave function $f(m)$ is defined on $[0, n/2 - 1]$. If $n/2 - 1 \leq \lfloor l \rfloor - 1$, then we have $\hat{S}_n \leq f(n/2 - 1)$; otherwise, m^* must be a value among $\{\lfloor l \rfloor - 1, \lfloor l \rfloor - 1/2, \lfloor l \rfloor, \lfloor l \rfloor + 1/2, \lfloor l \rfloor + 1\}$ that satisfies $m^* \leq n/2 - 1$ and achieves the maximum of $f(m)$. Together with Eq. (41) and the fact that the system time is nonnegative, we complete the proof. $\square$

By choosing appropriate variability parameters (Γ_a and Γ_s), we can accurately approximate the expected system time and thus the expected PAoI at each monitor, given that the expected interarrival time is already known (i.e., $1/\lambda$).

V. NUMERICAL RESULTS

In this section, we perform extensive simulations to evaluate the accuracy of our theoretical results. We first introduce

$(\theta_0, \theta_1, \theta_2)$	Single-Source	Two-Source
θ_0	-0.376	-1.302
θ_1	3.978	6.021
θ_2	0.5	0.7

TABLE I: Service adaptation regimes.

	Methods	Exponential	Normal	Uniform
Single-Source	Kingman's bound	33.86%	22.58%	14.89%
	Robust Approx. 1	32.01%	34.90%	36.49%
	Robust Approx. 2	8.32%	8.47%	9.28%
Two-Source	Robust Approx. 3	12.68%	10.05%	9.79%

TABLE II: Error percent of different approximation methods.

how to approximate the expected system time of the steady-state queueing networks with the worst-case system times proposed in Theorem 1 or Theorem 2. Then, for the single-source system, we show that our results in Theorem 1 can approximate the expected PAoI much better than the bounds in the literature (e.g., one bases on the Kingman's bound), especially in the light load case. In the end, we also show that the bound in Theorem 2 can also approximate the expected PAoI in two-source case very well.

A. Variability Parameters

Note that the bounds in Theorem 1 and Theorem 2 do not depend on the specific distribution of the interarrival time and service time. The update arrival process and service process are fully characterized by the primitive data (λ, σ_a^2) and (μ, σ_s^2), respectively, where σ_a^2 and σ_s^2 denote the variance of the interarrival time and service time, respectively. Therefore, it remains to translate the stochastic primitive data into uncertainty sets with appropriate variability parameters (Γ_a, Γ_s) such that the bounds proposed in Theorem 1 and Theorem 2 can approximate the expected system time of steady-state queueing networks well.

Inspired by the Kingman's bound[4] [15], a mapping function is provided in [14] that describes the variability parameters in terms of the distributions' first and second statistics,

$$\Gamma_a = \sigma_a, \Gamma_s = (\theta_0 + \theta_1 \sigma_s^2 + \theta_2 \sigma_a^2 \rho^2)^{1/2} - \sigma_a, \quad (46)$$

where $\rho = \lambda/\mu$ is the traffic density and $(\theta_0, \theta_1, \theta_2)$ are constants that can be derived from linear regression. Specifically, in order to obtain appropriate values of $(\theta_0, \theta_1, \theta_2)$, we first simulate multiple instances of the queue for various parameters of $(\rho, \sigma_a, \sigma_s)$ and different arrival and service distributions. Then we employ the linear regression to generate appropriate values for $(\theta_0, \theta_1, \theta_2)$ to adapt the value $\hat{S}_n$ obtained in Theorem 1 (or Theorem 2) to the expected value of the simulated system time. This allows us to build a dictionary or a look-up table of variability parameters values for given arrival and service distributions that makes the following approximation $\mathbb{E}[S(\mathbf{T}, \mathbf{X})] \approx \hat{S}_n(\Gamma_a, \Gamma_s)$. Table I provides the resulting $(\theta_0, \theta_1, \theta_2)$ for each adaption regimes.

[4]The Kingman's bound shows that the expected system time can be bounded by $\mathbb{E}[S] \leqslant \frac{\lambda}{2} \cdot \frac{\sigma_a^2 + \sigma_s^2}{1-\rho} + \frac{1}{\mu}$.

B. Single-Source Systems

We first consider the single-source setting, where the source generates updates with rate λ, and the service rate of the server is fixed with $\mu = 1$. We consider three different distributions for the interarrival times and service times: i) exponential distribution, ii) normal distribution, and iii) uniform distribution. For the normal distribution, we take the absolute value so that only positive interarrival time and service time are used. We consider the normal distribution and uniform distribution in order to show that our bound does not rely on the properties of specific distributions (e.g., the memoryless property of the exponential distribution).

In Fig. 4, we plot the PAoI as the arrival rate λ increases when the service rate $\mu = 1$. We name the approximation derived from [14] and Theorem 1 as *Robust Approximation 1* and *Robust Approximation 2*, respectively. Note that the variability parameters used in our Robust Approximation 2 are from Table I. Then, we compare the Kingman's bound, Robust Approximation 1 and Robust Approximation 2 against the simulated PAoI. Though the Kingman's bound and Robust Approximation 1 are originally defined for the expected system time, they can also be used as the PAoI bounds by simply adding the expected interarrival time $1/\lambda$. In order to demonstrate the detailed performance of different bounds, the error percent (which is the sum of the difference between the simulated PAoI and the bound normalized by the simulated PAoI in each arrival rate, divided by the total number of arrival rates) of each bound are also given in Table II. We run the simulations with a large number of arrivals ($n > 10^4$) to ensure that the steady state is reached. For each setting, we run the simulation for 50 rounds and take the average.

First, we can see that the PAoI is large under both light and high loads for all the considered distributions. This is because under the light load, the update arrival rate is low and the interarrival time between updates is large, which leads to a large PAoI; while under the high load, lots of updates wait to be served in the queue and the delay of those updates becomes large, which also leads to a high PAoI. This implies that we need to approximate PAoI well under both light and high loads. Second, we observe that Robust Approximation 2 we proposed can approximate the PAoI very well in all different settings. From Fig. 4b and Fig. 4c, we can see that the robust queueing approach is quite general and works well under normal and uniform distributions as well.

Compared to Robust Approximation 1, Robust Approximation 2 is much closer to the simulation results when the traffic load is light. This complies with our theoretical analysis, since Robust Approximation 2 cares about both light load and high load while Robust Approximation 1 is developed more towards approximating the system time under the high load. For the Kingman's bound, it is known that the result is tight with light-tailed distribution in high load. However, similar to Robust Approximation 1, the approximation of Kingman's bound is also loose in the light load. Therefore, our proposed bound provides a competitive alternative to approximating the PAoI
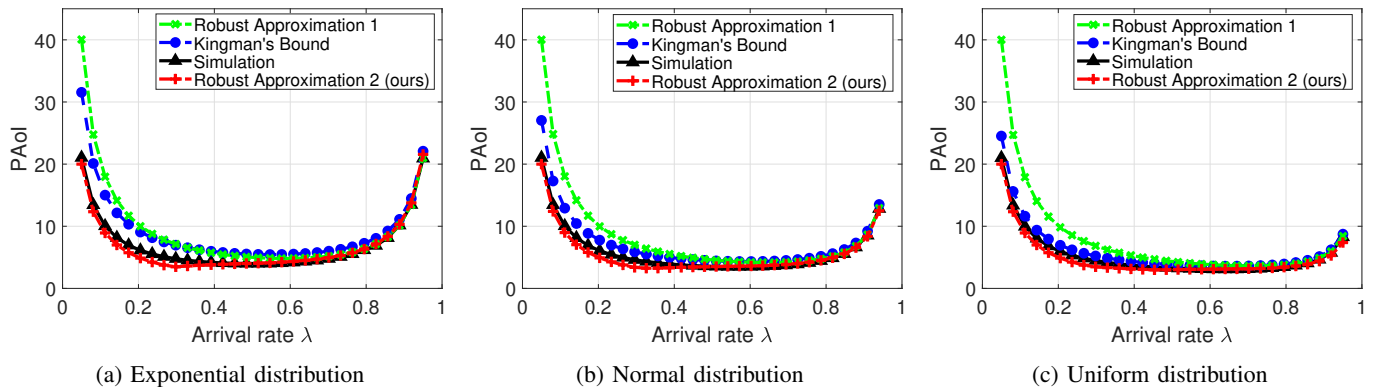
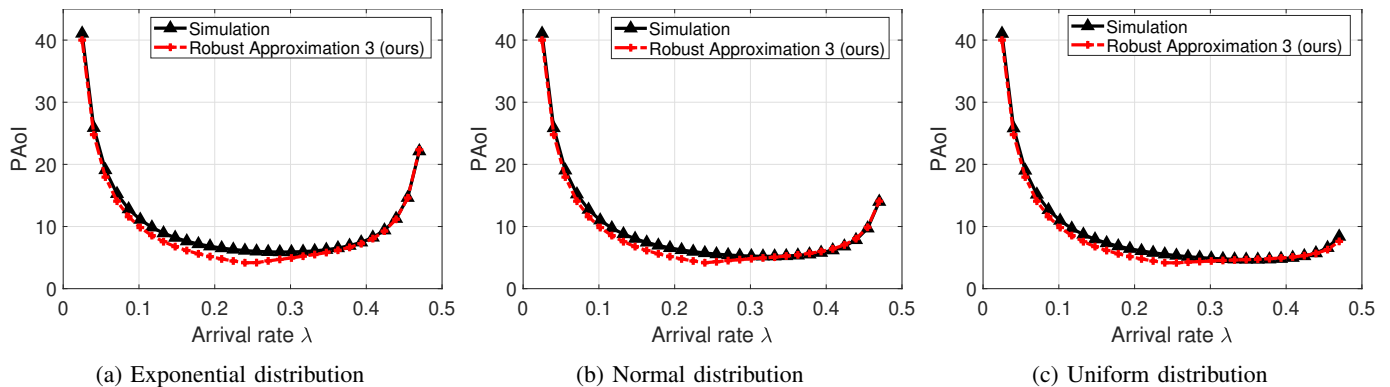Fig. 4: Peak AoI under different distributions in the single-source setting.

Fig. 5: Peak AoI under different distributions in the two-source setting with symmetric arrivals.

performance in information-update systems.

C. Two-Source Systems

In this subsection, we consider the information-updating system with two symmetric sources and one single server. Here the two sources generate updates with the same interarrival time distribution (also the same expected interarrival time $1/\lambda$) and the same service time distribution (also the same expected service time $1/\mu$, i.e., they are symmetric. Similar to the single-source case, we also consider three different distributions, where the interarrival times and service times are both exponential distribution, normal distribution and uniform distribution, respectively. We name the approximation derived from Theorem 2 as *Robust Approximation 3*, which adopts the variability parameters in Table I. With increasing arrival rate λ and fixed service rate $\mu = 1$, the PAoI performances of one arbitrary source are shown in Fig. 5 and Table II.

Note that the Kingman's bound and Robust Approximation 1 cannot be used to approximate the PAoI in the two-source setting since they are only derived for the single-source systems. Again, we observe that Robust Approximation 3 can approximate the average PAoI well under both light and high traffic loads.

VI. CONCLUSION

In this paper, we applied the robust queueing theory to analyzing the AoI performance in the communication systems.

By modeling the uncertainty in the stochastic arrival and service processes using uncertainty sets, we provided a robust bound of the worse-case delay that can be used to approximate the expected PAoI in the single-source single-server systems. Furthermore, we generalized our bound to the two-source single-server systems with symmetric arrivals. We showed that our bounds work well under both light and high loads and outperform prior bounds (e.g., the Kingman's bound), which could be quite inaccurate under the light load. Moreover, our results do not rely on the property of specific distributions and can thus be widely applied to more general settings.

There are several of interesting questions that are worth future work. For example, we only show that the robust queueing bound works well for the light-tailed systems in the simulations. However, the robust queueing theory is also capable of modeling the heavy-tailed behaviors. It would be interesting to investigate the effectiveness of the proposed bounds in the heavy-tailed systems. In addition, we assume that the service time and interarrival time have the same tail coefficient. It would be nice to generalize the results for the case where the service time and interarrival time have different tail coefficients. Finally, it is also worth studying the impact of the dependence between the arrival and service process [27].

REFERENCES

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2011, pp. 350–358.

[2] S. Kaul, R. Yates, and M. Gruteser, "On piggybacking in vehicular networks," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011, pp. 1–5.

[3] ——, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2731–2735.

[4] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 1583–1587.

[5] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1–6.

[6] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[7] Y. Sang, B. Li, and B. Ji, "The power of waiting for more than one response in minimizing the age-of-information," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[8] N. Pappas, J. Gunnarsson, L. Kratz, M. Kountouris, and V. Angelakis, "Age of information of multiple sources with queue management," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 5935–5940.

[9] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *2012 IEEE International Symposium on Information Theory Proceedings*, 2012, pp. 2666–2670.

[10] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1681–1685.

[11] O. Dogan and N. Akar, "The Multi-Source Preemptive M/PH/1/1 Queue with Packet Errors: Exact Distribution of the Age of Information and Its Peak," *arXiv e-prints*, p. arXiv:2007.11656, Jul. 2020.

[12] M. Moltafet, M. Leinonen, and M. Codreanu, "An approximate expression for the average aoi in a multi-source m/g/1 queueing model," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[13] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011. [Online]. Available: https://doi.org/10.1137/080734510

[14] C. Bandi, D. Bertsimas, and N. Youssef, "Robust queueing theory," *Operations Research*, vol. 63, no. 3, pp. 676–700, 2015.

[15] J. F. Kingman, "Inequalities in the theory of queues," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 32, no. 1, pp. 102–110, 1970.

[16] F. Ciucu and F. Poloczek, "Two extensions of kingman's gi/g/1 bound," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 2, no. 3, pp. 1–33, 2018.

[17] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[18] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of information: A new metric for information freshness*. Morgan & Claypool Publishers, 2019.

[19] R. D. Yates, Y. Sun, I. Brown, D. Richard, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of Information: An Introduction and Survey," *arXiv e-prints*, p. arXiv:2007.08564, Jul. 2020.

[20] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 2569–2573.

[21] F. Li, Y. Sang, Z. Liu, B. Li, H. Wu, and B. Ji, "Waiting but not Aging: Optimizing Information Freshness Under the Pull Model," *arXiv e-prints*, p. arXiv:1912.08722, Dec. 2019.

[22] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8305–8324, 2019.

[23] J. P. Champati, H. Al-Zubaidy, and J. Gross, "On the distribution of aoi for the gi/gi/1/1 and gi/gi/1/2* systems: Exact expressions and bounds," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 37–45.

[24] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

[25] C. Bandi and D. Bertsimas, "Tractable stochastic analysis in high dimensions via robust optimization," *Mathematical programming*, pp. 1–48, 2012.

[26] D. V. Lindley, "The theory of queues with a single server," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, no. 2. Cambridge University Press, 1952, pp. 277–289.

[27] W. Whitt and W. You, "Using robust queueing to expose the impact of dependence in single-server queues," *Operations Research*, vol. 66, no. 1, pp. 184–199, 2018.