Statistical Guarantee Optimization for Age of Information for the D/G/1 Queue

Jaya Prakash Champati*, Hussein Al-Zubaidy†, James Gross*

*School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

†School of Information Technology (ITE), Halmstad University, Halmstad, Sweden

Abstract—**Age of Information (AoI) has proven to be a useful metric in networked systems where timely information updates are of importance. Recently, minimizing the "average age" has received considerable attention. However, various applications pose stricter age requirements on the updates which demand knowledge of the AoI distribution. In this work, we study the distribution of the AoI and devise a problem of minimizing the tail of the AoI distribution function with respect to the frequency of generating information updates, i.e., the sampling rate of monitoring a process, for the D/G/1 queue model under FCFS queuing discipline. We argue that computing an exact expression for the AoI distribution may not always be feasible. Therefore, we opt for computing a bound on the tail of the AoI distribution and use it to formulate a tractable α-relaxed Upper Bound Minimization Problem (α-UBMP), where $\alpha > 1$ is an approximation factor. This approximation can be used to obtain "good" heuristic solutions. We demonstrate the efficacy of our approach by solving α-UBMP for the D/M/1 queue. We show, using simulation, that the rate solutions obtained are near optimal for minimizing the tail of the AoI distribution.**

I. INTRODUCTION

Providing real-time status updates is a key requirement for many networked systems which support emerging time-critical applications such as autonomous vehicle systems, automation of manufacturing processes, sending phasor data updates to a power grid system etc. In contrast to human-related data applications, the freshness of the status updates is critical for optimal control in these networked systems. In this respect, the traditional goal of maximizing throughput becomes irrelevant as the freshness of the status updates not only depends on queueing and transmission delays in the network, but also on the frequency of generating updates at the source. Recently, the Age of Information (AoI), proposed in [1], has emerged as a relevant metric in quantifying the freshness of the status updates. It is defined as the time elapsed since the generation of the latest status update received at the destination.

Since timely updates of status information is essential for optimal control in the above systems, providing QoS guarantees for AoI is of prime importance. Interestingly, in contrast to *delay* which increases with update sampling rate, AoI has the property that it increases at both high and low sampling rates [2]. This property of AoI inspired several theoretical works focusing on minimizing some function of AoI for simple yet important system models. Following this line of research, we consider a source-destination pair, where the source samples a process of interest, and generates status updates (packets) at a constant rate R. These packets are queued in an infinite buffer and are served in a First-Come-First-Serve (FCFS) fashion. We model this system using a D/G/1 queuing system and study the problem of finding an optimal sampling rate that minimizes the AoI violation probability for a given *age limit d*.

In [2], the authors have addressed the minimization of the *time-average age* for M/M/1, M/D/1 and D/M/1 queuing systems. In [3], the authors optimized AoI for the M/M/1 queueing system with multiple sources. Optimal generation rate was also studied in [4], [5] for a source that is constrained by a time-varying energy process. In [6], the authors have studied the problem of computing optimal arrival rate for minimizing *average peak age*, proposed in [7], for a multi-class M/G/1 system and proposed an approximate solution. In contrast to previous works, the authors in [8], [9] considered a general class of non-decreasing functions of AoI. While optimal updating policies are studied in [8], optimal scheduling policies in multi-hop networks are studied in [9]. In contrast to the previous works, the authors in [10] have studied the relation between the distributions of AoI, peak AoI and the sojourn time, under FCFS discipline. However, they have not presented a standalone expression for the distribution of AoI and did not study the optimization of the sampling rate.

To the best of our knowledge, existing theoretical results on optimizing AoI have focused mainly on some form of "average age" function. In contrast, here we consider the AoI distribution, more precisely, we evaluate the probability that a given age limit d is violated. This metric represents for instance a stochastic guarantee on the timeliness of the state information regarding the process under observation. It can thus be offered by a network as a QoS requirement to the application. However, computing an exact expression for the AoI violation probability is not straightforward. Therefore, we resort to working with tractable upper bounds which facilitate the computation of heuristic solutions.

The main contributions of this work are summarized below:

- We characterize, for the first time, the AoI distribution for the D/G/1 queue. More precisely, we compute the probability that a given age limit d is violated.
- We formulate the AoI violation probability minimization problem $\mathcal{P}$, and show that it is equivalent to minimizing the violation probability of the departure instant of a certain packet over the rate region $[\frac{1}{d}, \mu)$, where μ is the service rate. This in turn enables us to propose an

Upper Bound Minimization Problem (UBMP) which can be used in finding a "good" heuristic solution for $\mathcal{P}$.

- Noting that solving UBMP involves computation of a sum of infinite terms which can be intractable, we propose α-relaxed UBMP (α-UBMP) the solution of which has $\alpha > 1$ approximation ratio with respect to UBMP.
- We demonstrate the usefulness of α-UBMP by solving it for the D/M/1 queue and show using simulation that the rate solutions obtained are near optimal for $\mathcal{P}$.

II. SYSTEM MODEL AND PROBLEM STATEMENT

We consider a D/G/1 queuing system under FCFS scheduling. A source generates packets (status updates) at a constant rate R. Thus, R models the sampling rate of a process under observation. Let $T = \frac{1}{R}$ denote the inter-arrival time between any two packets. We index the packets by $n \in \{0, 1, 2 \ldots\}$. Let $T_A(n, R) = \frac{n}{R}$ denote the arrival instant of packet n and $T_D(n, R)$ the corresponding departure instant. The service time for packet n is given by a random variable X_n, where $\{X_n, n \geq 0\}$ are i.i.d. with mean service rate $\mu = \frac{1}{\mathbb{E}[X_n]}$.

At the destination, we are interested in maintaining timely state information of the process. We are thus interested in the AoI metric, denoted by $\Delta(t, R)$, and is defined as:

$$\Delta(t, R) \triangleq t - \max\{T_A(n, R) : T_D(n, R) \leq t\}. \quad (1)$$

For a given *age limit* requirement $d > 0$, in the following we study the distribution of AoI by characterizing its violation probability, i.e., $\mathbb{P}(\Delta(t, R) > d)$, both in the transient and the steady state of the D/G/1 queue. Given the stochastic characterization, we are furthermore interested in finding the sampling rate R that minimizes the AoI violation probability. Given d, we are thus interested in solving the following problem $\mathcal{P}$:

$$\min_R V(d, R),$$

where $V(d, R) \triangleq \lim_{t \to \infty} \mathbb{P}(\Delta(t, R) > d)$ is the steady state AoI violation probability. Let $R^*(d)$ and $V(d, R^*)$ denote the optimal rate and the optimal value, respectively, for $\mathcal{P}$.

Henceforth, we drop R from the notation when it is obvious from the context, for the sake of notation simplicity.

III. AOI VIOLATION PROBABILITY ANALYSIS

In this section, we study the distribution of the AoI for the model described above. We start by investigating structural characteristics of the stochastic behavior of AoI. In the following, we rehearse the max-plus representation of Reich's equation to model the evolution of the considered D/G/1 system. For any realization of the service times, the relation between $T_D(n, R)$, $T_A(n, R)$ and $\{X_n\}$, is given by [11]:

$$T_D(n, R) = \max_{0 \leq v \leq n} \left\{ T_A(n - v, R) + \sum_{i=0}^{v} X_{n-i} \right\}. \quad (2)$$

Consider the definition in (1), for $\Delta(t, R)$ not to exceed the age limit d, the latest departure at $t \geq 0$ must have arrived no earlier than $[t - d]^+$, since in our model, no arrival is allowed before $t = 0$. Therefore, to study the distribution of $\Delta(t, R)$,
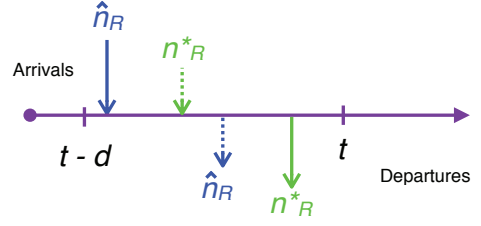

Fig. 1: Time-line of events for Case 2 in Lemma 1 proof.

we tag the packet arriving on or immediately after $t - d$ and use it to characterize this process. Given rate R, let $\hat{n}_R$ denote the first arrival on or immediately after time $t - d$, given by

$$\hat{n}_R \triangleq \lceil R(t - d) \rceil. \quad (3)$$

In the following lemma we present a key insight regarding the transient characterization of the AoI violation probability.

Lemma 1. *Given a D/G/1 queuing system with arrival rate R, age limit d, and $t < \infty$, if there exists n such that $t - d \leq \frac{n}{R} < t$, then $\mathbb{P}\{\Delta(t, R) > d\} = \mathbb{P}\{T_D(\hat{n}_R) > t\}$, otherwise, $\mathbb{P}\{\Delta(t, R) > d\} = 1$.*

Proof. Let n_R^* is the latest packet departure at t, i.e., $n_R^* = \arg\max_n\{T_D(n, R) \leq t\}$. Thus, $\Delta(t, R) = t - T_A(n_R^*)$.

Case 1: If an n such that $t - d \leq \frac{n}{R} < t$ does not exist, i.e., there is no arrival during the time interval $[t - d, t)$, then the arrival time of n_R^* must be strictly less than $t - d$, i.e., $T_A(n_R^*) < t - d$. Therefore,

$$P(\Delta(t, R) > d) = P(t - T_A(n_R^*) > d) = 1.$$

Case 2: If there exists n such that $t - d \leq \frac{n}{R} < t$, then $t - d \leq \frac{\hat{n}_R}{R} < t$, since $\hat{n}_R$ is the first arrival on or after time $t - d$, see Figure 1. In this case, we show that the event $\{\Delta(t, R) \leq d\}$ is equivalent to the event $\{T_D(\hat{n}_R) \leq t\}$. Suppose that the event $\{\Delta(t, R) \leq d\}$ occurred, then $T_A(n_R^*) \geq t - d$. By definition of $\hat{n}_R$, we should have $T_A(\hat{n}_R) \leq T_A(n_R^*)$ which implies $T_D(\hat{n}_R) \leq T_D(n_R^*) \leq t$, due to FCFS assumption. Therefore, $\{\Delta(t, R) \leq d\} \subseteq \{T_D(\hat{n}_R) \leq t\}$.

To prove equivalence of the two events, we show that the relation above also holds the other way around. Suppose that the event $\{T_D(\hat{n}_R) \leq t\}$ occurred. Again, it should be true that $T_A(n_R^*) \geq T_A(\hat{n}_R)$. Otherwise, $T_D(n_R^*) < T_D(\hat{n}_R) \leq t$ which contradicts the definition of n_R^*. Therefore,

$$\Delta(t, R) = t - T_A(n_R^*) \leq t - T_A(\hat{n}_R) \leq t - (t - d) = d.$$

This implies that $\{T_D(\hat{n}_R) \leq t\} \subseteq \{\Delta(t, R) \leq d\}$. Therefore, the equivalence holds and the result is proven. □

Note that **Case 1** in the above proof essentially represents an undersampling of the process under observation, i.e., with respect to the choice of d at the current point t the sampling rate R is simply too low to maintain the target age limit.

We next present the steady state results for the D/G/1 system based on the results obtained in Lemma 1.

Theorem 1. *Given age limit d, the steady state distribution of AoI for the D/G/1 queue is characterized as follows:*

1) *If $R \geq \frac{1}{d}$, then*

$$\lim_{t \to \infty} \mathbb{P}\{\Delta(t, R) > d\} = \lim_{t \to \infty} \mathbb{P}\{T_D(\hat{n}_R) > t\}. \quad (4)$$

2) *Else if $R < \frac{1}{d}$, then*

$$\limsup_{t \to \infty} \mathbb{P}\{\Delta(t, R) > d\} = 1,$$
$$\liminf_{t \to \infty} \mathbb{P}\{\Delta(t, R) > d\} = \lim_{t \to \infty} \mathbb{P}\{T_D(\hat{n}_R) > t\}.$$

Proof. For the two cases above consider the following:

Case 1 ($R \geq \frac{1}{d}$): Since the samples are generated at a constant rate, for $R \geq \frac{1}{d}$ we claim that there exist an n such that $t - d \leq \frac{n}{R} < t$, for all t. We first prove this claim for $R > \frac{1}{d}$. We have

$$T_A(\hat{n}_R) = \frac{\lceil R(t-d) \rceil}{R} \leq \frac{R(t-d) + 1}{R} < t.$$

Furthermore, since $t - d \leq T_A(\hat{n}_R)$ for any t by definition, the claim holds at least for $\hat{n}_R$, for $R > \frac{1}{d}$. To prove the claim for $R = \frac{1}{d}$, we consider

$$t - d \leq \frac{n}{R} < t \Leftrightarrow \frac{n}{R} < t \leq d + \frac{n}{R} \Leftrightarrow n < Rt \leq n + 1.$$

Note that for any R and t there always exists an n such that the last inequality above holds. Therefore, the claim is true and Case 1 follows from Lemma 1 by letting t go to infinity.

Case 2 ($R < \frac{1}{d}$): In this case, the existence of n such that $t - d \leq \frac{n}{R} < t$ depends on t. Again, using Lemma 1, for all t where this condition is satisfied we have $\mathbb{P}\{\Delta(t, R) > d\} = \mathbb{P}\{T_D(\hat{n}_R) > t\}$. For all other values of t, we have $\mathbb{P}\{\Delta(t, R) > d\} = 1$. This implies that as t goes to infinity the violation probability oscillates between $\mathbb{P}\{T_D(\hat{n}_R) > t\}$ and 1. Thus, we obtain the limit supremum and the limit infimum. $\quad\square$

Intuitively, given R, the support of the steady state AoI distribution should be $[\frac{1}{R}, \infty)$, because AoI cannot be less than $\frac{1}{R}$ when the samples are generated at rate R. Not only Theorem 1 asserts this intuitive reasoning, but also characterizes the limit infimum and limit supremum for the region $d < \frac{1}{R}$, where the AoI violation probability does not exist. Therefore, to ensure the existence of the AoI violation probability we consider the feasible rate region $[\frac{1}{d}, \mu)$, where $R < \mu$ ensures queue stability. In light of this, and using (4) from Theorem 1, we formulate an equivalent problem $\tilde{\mathcal{P}}$ as follows:

$$\begin{aligned} \min_{R} \quad & \lim_{t \to \infty} \mathbb{P}(T_D(\hat{n}_R) > t), \\ \text{s.t.} \quad & \frac{1}{d} \leq R < \mu. \end{aligned} \quad (5)$$

We next present the upper bound minimization problems.

IV. UPPER BOUND MINIMIZATION

In order to find a solution for $\tilde{\mathcal{P}}$, we must first evaluate the probability $\mathbb{P}\{T_D(\hat{n}_R) > t\}$, where $T_D(n)$ is given by (2). Note that $T_D(n)$ is random, since the service process $\{X_n, n \geq 0\}$ is random, and is given in terms of the maximum of $n+1$ random variables. Hence, obtaining an exact expression is tedious. Therefore, we opt for a more tractable

approach by using probabilistic inequalities to obtain bounds on the distribution of $T_D(\hat{n}_R)$. Consequently, we propose the Upper Bound Minimization Problem (UBMP) and its more computationally efficient counterpart α-relaxed UBMP (α-UBMP) to obtain a near optimal solution for $\tilde{\mathcal{P}}$, where the solution of α-UBMP has α-approximation ratio with respect to UBMP.

A. A Bound for the Distribution of T_D

As mentioned earlier, the evaluation of the distribution function of $T_D(n)$ requires the computation of the distribution of the maximum of random variables. Fortunately, there are several approaches that have been used in the literature to estimate this probability. One such approach approximates the probability of the maximum by the maximum probability, i.e., $\mathbb{P}\{\max_i Y_i > y\} \approx \max_i \mathbb{P}\{Y_i > y\}$. However, this approximation is not always accurate and in some cases may result in very large deviation from the actual distribution. Hence, it cannot be used when reliability of the solution must be well defined as it is the case here. An alternative approach is to use extreme value theorem. However, the obtained extreme value distributions are not always tractable. A more promising approach is to use Boole's inequality, commonly known as the "union bound," where the probability of a union of events is bounded by the sum of their probabilities. The bound obtained in our case is not only tractable, but is also arguably tight. In the following lemma, we present this upper bound for the distribution function $\lim_{t \to \infty} \mathbb{P}\{T_D(\hat{n}_R) > t\}$.

Lemma 2. *Given d and the corresponding $\hat{n}_R$ as defined in (3), then*

$$\lim_{t \to \infty} \mathbb{P}(T_D(\hat{n}_R) > t) \leq \sum_{v=0}^{\infty} \Phi(v, R),$$

where

$$\Phi(v, R) \triangleq \mathbb{P}\left\{ \sum_{i=0}^{v} X_{\hat{n}_R - i} > d + \frac{v-1}{R} \right\}. \quad (6)$$

Proof. Using (2), we have

$$\mathbb{P}\{T_D(\hat{n}_R) > t\} = \mathbb{P}\left\{ \max_{0 \leq v \leq \hat{n}_R} \left(T_A(\hat{n}_R - v) + \sum_{i=0}^{v} X_{\hat{n}_R - i} \right) > t \right\}$$

$$\leq \sum_{v=0}^{\hat{n}_R} \mathbb{P}\left\{ \sum_{i=0}^{v} X_{\hat{n}_R - i} > t - \frac{\hat{n}_R - v}{R} \right\}$$

$$\leq \sum_{v=0}^{\hat{n}_R} \mathbb{P}\left\{ \sum_{i=0}^{v} X_{\hat{n}_R - i} > t - \frac{R(t-d) + 1 - v}{R} \right\}$$

$$= \sum_{v=0}^{\hat{n}_R} \underbrace{\mathbb{P}\left\{ \sum_{i=0}^{v} X_{\hat{n}_R - i} > d + \frac{v-1}{R} \right\}}_{\triangleq \Phi(v, R)}.$$

Above we first applied the union bound (step 2) and then used $\hat{n}_R = \lceil R(t-d) \rceil \leq R(t-d) + 1$ (step 3). The result follows by noting that $\hat{n}_R$ goes to infinity as t goes to infinity. $\quad\square$

B. UBMP Formulation

Using (5) and Lemma 2 we propose the following UBMP problem.

$$\min_R \quad \sum_{v=0}^{\infty} \Phi(v, R)$$
$$\text{s.t.} \quad \frac{1}{d} \le R < \mu. \tag{7}$$

It is worth noting that the function $\Phi(0, R)$ is non-increasing in R while the functions $\{\Phi(v, R) : v > 1\}$ are non-decreasing in R. This indicates that, for any service time distribution, the objective function of UBMP will have a global minimum for R in the interval $[\frac{1}{d}, \mu)$.

A shortcoming of UBMP is that its objective function involves computation of sum of infinite terms and each term requires computation of the distribution of sum of service times. We will see later, even for the D/M/1 queue computing these terms quickly become intractable as v increases. To this end, we resort to solving α-UBMP which is described next.

C. α-UBMP Formulation

For any $\alpha > 1$, α-UBMP computes only the first K_α terms of $\sum_{v=0}^{\infty} \Phi(v, R)$, where K_α is chosen such that the resulting sum is at most α times $\sum_{v=0}^{\infty} \Phi(v, R)$. Given R, d, μ and α, parameter K_α is defined as the minimum K that satisfies

$$\sum_{v=0}^{\infty} \Phi(v, R) \le \alpha \sum_{v=0}^{K-1} \Phi(v, R). \tag{8}$$

Then, α-UBMP is defined as follows:

$$\min_R \quad \sum_{v=0}^{K_\alpha - 1} \Phi(v, R)$$
$$\text{s.t.} \quad \frac{1}{d} \le R < \mu. \tag{9}$$

We refer to $\sum_{v=0}^{K_\alpha - 1} \Phi(v, R)$ as the α-*relaxed upper bound*. Let $\hat{R}_\alpha(d)$ denote an optimal solution for α-UBMP.

To compute such K_α we first find an upper bound for $\sum_{v=K}^{\infty} \Phi(v, R)$ which is accomplished by using Chernoff's bound. We demonstrate this by solving α-UBMP for the D/M/1 queue in the following section.

V. Rate Optimization for The D/M/1 queue

In this section, we consider a system with exponentially distributed service time with rate μ, i.e., $X_n \sim \exp(\mu)$. This results in the D/M/1 queuing model. In this special case, $\Phi(v, R)$ is the distribution function of a sum of $v + 1$ i.i.d. exponential random variables, which is given by a gamma distribution with scale parameter $v + 1$ and rate parameter μ. Therefore,

$$\Phi(v, R) = \frac{\Gamma(v + 1, \mu(d + (v - 1)/R))}{v!}, \tag{10}$$

where $\Gamma(x, a)$ is the upper incomplete gamma function:

$$\Gamma(x, a) = \int_a^{\infty} y^{x-1} e^{-y} dy.$$

Interestingly, UBMP is a convex optimization problem for the D/M/1 queue which is stated in the following theorem.

Theorem 2. *For the D/M/1 queue, the UBMP is a convex optimization problem with respect to $\frac{1}{R}$.*

Note that $\Phi(1, R)$ is insensitive to R. For all other $v \ne 1$, we prove that $\Phi(v, R)$ given in (10) is strictly convex with respect to $\frac{1}{R}$. The complete proof is given in Appendix A.

Despite proving the convexity of UBMP, we face the following issues in solving it. We are unable to compute the optimal R using the KKT conditions as it involves solving a set of complex equations containing sums of infinite terms. We may use efficient convex optimization algorithms. However, this requires computing $\Phi(v, R)$ for all v values, and both the upper incomplete gamma function and $v!$ become computationally prohibitive as v becomes large. Thus, we resort to solving α-UBMP approximation that was formulated in (9).

As noted before, α-UBMP requires computing the parameter K_α. To do that, we need to find the minimum K that satisfies (8). To this end, we first present an upper bound for $\sum_{v=K}^{\infty} \Phi(v, R)$ which is stated in the following lemma:

Lemma 3. *For $K \ge 0$, we have*

$$\sum_{v=K}^{\infty} \Phi(v, R) \le \min_{0 < s < \mu} \frac{e^{-s(d + (K-1)/R)}}{(1 - s/\mu)^K (1 - s/\mu - e^{-s/R})}. \tag{11}$$

The proof is given in Appendix B.

The following corollary characterizes K_α for the D/M/1 system using the bound given by Lemma 3.

Corollary 1. *Given R, d, μ, and α, K_α is the minimum K value that satisfies*

$$\sum_{v=0}^{K-1} \Phi(v, R) \ge \min_{0 < s < \mu} \frac{1}{\alpha - 1} \cdot \frac{e^{-s(d + (K-1)/R)}}{(1 - s/\mu)^K (1 - s/\mu - e^{-s/R})}. \tag{12}$$

The proof of the corollary is given in Appendix C.

Given α, we compute K_α numerically using the condition in Corollary 1. Since K_α varies with R, d and μ, we simply choose the maximum K_α that satisfies (12) for the given set of R, d and μ. From Theorem 2, we infer that α-UBMP is also a convex optimization problem with respect to $\frac{1}{R}$. Therefore, we use Newton's method to compute the optimal $\hat{R}_\alpha(d)$.

VI. Numerical Evaluation

In this section, we evaluate the performance of α-UBMP solutions for the D/M/1 queue. We first study the tightness of the α-relaxed upper bound in comparison to the AoI violation probability $V(d, R)$ obtained from simulations. We then evaluate the quality of numerically computed solution $\hat{R}_\alpha(d)$ by comparing the simulation-based estimates $V(d, \hat{R}_\alpha)$ against $V(d, R^*)$. The numerics are done by MATLAB while the simulations are implemented in C. We use the following default parameters: $\mu = 1.5$, $R = 1$, $d = 10$ and $\alpha = 1.01$.
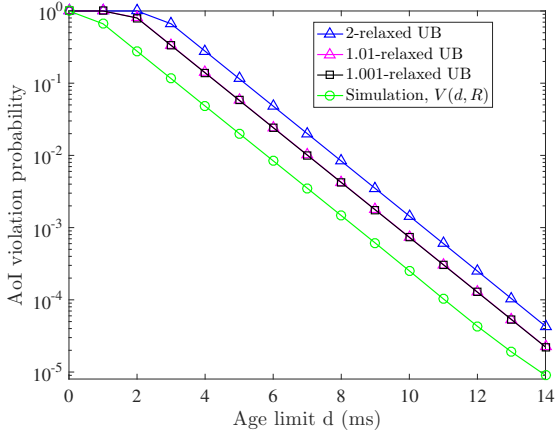
Fig. 2: Tightness of the α-relaxed upper bound with varying d. $\mu = 1.5$ packets/ms, $R = 1$ packet/ms, $K_{1.001} = 143$, $K_{1.01} = 118$ and $K_2 = 66$.

1) Tightness of the α-Relaxed Upper Bound: In Figure 2, we present different α-relaxed upper bounds in comparison with the AoI violation probability by varying age limit d. We observe that the bounds are within an order of magnitude and become tighter as α approaches 1. Another important observation is that the decay rates of the upper bounds are equal and follow the decay rate of $V(d, R)$. We note that the corresponding K_α values are chosen such that condition (12) is satisfied for all $d \in \{2, 4, 6, 8, 10\}$, $R = 1$ and $\mu = 1.5$.

In Figure 3, we compare the 5-relaxed upper bound with $V(d, R)$ by varying R. Clearly, both the upper bound and $V(d, R)$ have convex nature and a global minimum in the interval $[\frac{1}{d}, \mu)$. Observe that the upper bound follows the trend of $V(d, R)$ and only deviates at high sampling rates. This is an interesting property as it suggests that a rate that minimizes the upper bound also minimizes $V(d, R)$. Here $K_5 = 170$, which is again the value that satisfies (12) for all $R \in \{.2, .3, \ldots, 1.3\}$, $d = 10$ and $\mu = 1.5$. We couldn't present results for $\alpha < 5$, because this requires the use of larger K_α values and MATLAB couldn't compute the α-relaxed upper bound. The K_α values here are large due to the R value 1.3 which is close to μ. Fortunately, below we show that computing $\hat{R}_\alpha(d)$ does not require large K_α values as it turns out the value of $\hat{R}_\alpha(d)$ is typically not close to μ.

2) Quality of Heuristic Solution $\hat{R}_\alpha(d)$: In Figure 4, we compare $V(d, \hat{R}_\alpha)$ with $V(d, R^*)$ for different α values and varying age limit d. Note that the difference between $V(d, \hat{R}_\alpha)$ and $V(d, R^*)$ is negligible. This suggests that the solutions of α-UBMP are near optimal for $\mathcal{P}$. Also, we note that solving α-UBMP for $\alpha = 1.01$ and $\alpha = 1.25$ results in exact same solution. This can be attributed to the equal decay rate of α-relaxed upper bounds for different α as observed in Figure 2. This further suggests that we do not need to solve α-UBMP for α close to 1 in order to find a good heuristic solution. Also note that the values $K_{1.01} = 28$ and $K_{1.25} = 15$ are smaller compared to K_5 value for Figure 3, because these are
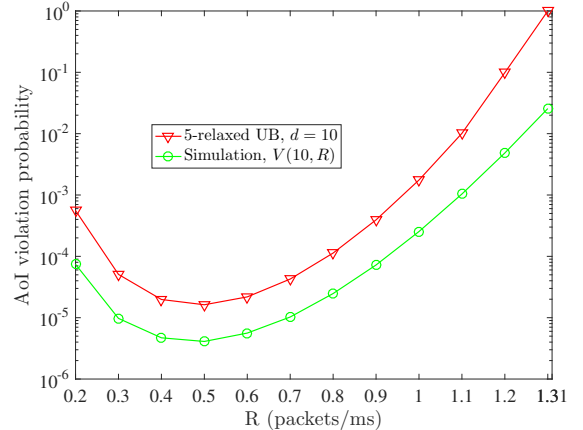

Fig. 3: Tightness of the α-relaxed upper bound with varying R. $\mu = 1.5$ packets/ms, $K_5 = 170$ and $d = 10$ ms.
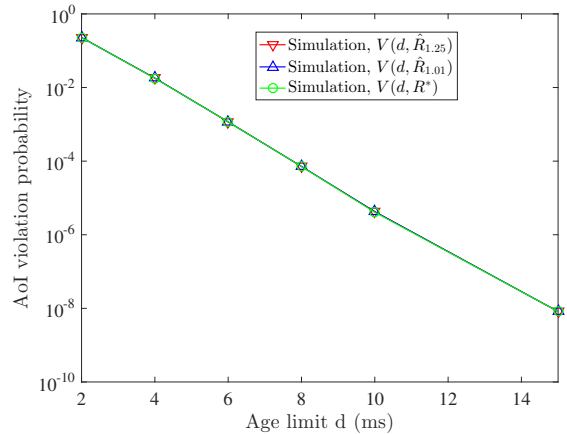

Fig. 4: Quality of α-UBMP solutions for varying d. $\mu = 1.5$ packets/ms, $K_{1.01} = 28$ and $K_{1.25} = 15$.

evaluated at $\hat{R}_\alpha(d)$ which does not fall near μ.

Finally, in Figure 5, we compare $V(d, \hat{R}_\alpha)$ with $V(d, R^*)$ for different age limits and varying service rate μ. Again, we notice negligible difference between the two. Therefore, we conclude that solutions to α-UBMP are near optimal for $\mathcal{P}$ for the D/M/1 queue.

We note that unlike the time-average age objective, which is minimized at $.515$ utilization factor for the D/M/1 queue [2], $R^*(d)$ and in turn the utilization factor that minimizes AoI violation probability depends on age limit d. For a comparison, in Figure 4 the optimal rates are found to be $\{.7, .6, .567, .525, .5, .267\}$ and the corresponding utilization factors are $\{.46, .4, .378, .35, .33, .178\}$.

VII. CONCLUSION AND FUTURE WORK

We characterize the distribution of the age of information (AoI) for the D/G/1 queue and discovered some of its important structural properties. We then use this knowledge to formulate an optimization problem $\mathcal{P}$ to find the optimal
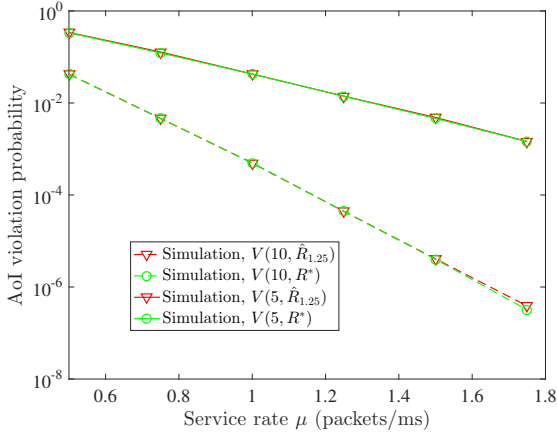
Fig. 5: Quality of α-UBMP solutions for varying μ. $d = 10$ ms and $K_{1.25} = 15$.

information update rate which minimizes the probability that AoI violates a given age limit d. Furthermore, we show that $\mathcal{P}$ is equivalent to the problem of minimizing the violation probability of the departure time of a tagged arrival $\hat{n}_R$ over the rate region $[\frac{1}{d}, \mu)$. Noting that finding an exact expression for the distribution of the departure time process may not be possible, we propose an upper bound minimization problem, UBMP, and its more computationally tractable version α-UBMP, which results in heuristic solutions. We solve α-UBMP for the D/M/1 queue and the numerical results suggest that the rate solutions of α-UBMP are near optimal for $\mathcal{P}$.

For future work, we are investigating the extension of our results to stochastic arrivals. We are also studying the computational complexity for solving α-UBMP and investigating more efficient solution methods, i.e., by identifying the range of α for which a good heuristic solution for $\mathcal{P}$ can be obtained.

References

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE SECON*, 2011.

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

[3] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE ISIT*, 2012.

[4] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE ISIT*, 2015.

[5] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. Information Theory and Applications Workshop (ITA)*, 2015.

[6] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE ISIT*, 2015.

[7] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE ISIT*, 2014.

[8] 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, Nov 2017.

[9] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, 2017.

[10] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *Proc. IEEE ISIT*, 2017.

[11] J. Liebeherr, "Duality of the max-plus and min-plus network calculus," *Found. and Trends in Networking*, vol. 11, no. 3-4, pp. 139–282, 2017.

Appendix

A. Proof of Theorem 2

Recall that $T = \frac{1}{R}$. We prove that $\frac{\partial^2 \Phi(v,T)}{\partial T^2} > 0$ for $v \neq 1$. Let $z = d + T(v - 1)$. We have

$$\frac{\partial \Phi(v,T)}{\partial T} = -\mu^{v+1}(v-1)z^v e^{-\mu z}/v!$$

$$\Rightarrow \frac{\partial^2 \Phi(v,T)}{\partial T^2} = -\mu^{v+1}(v-1)^2 z^{v-1} e^{-\mu z}(v - \mu z)/v!.$$

To prove the result it is sufficient to prove that $v - \mu z < 0$. Since $\frac{\mu}{R} = \mu T > 1$ and $d > \frac{1}{R} = T$, we have

$$v - \mu z = v(1 - \mu T) - \mu(d - T) < 0.$$

B. Proof of Lemma 3

From (6), we have

$$\Phi(v,R) \leq \mathbb{P}\left\{\sum_{i=0}^{v} X_{n-i} > d + \frac{v-1}{R}\right\}$$

$$\leq \min_{s>0} e^{-s(d+(v-1)/R)} \mathbb{E}[e^{s(\sum_{i=0}^{v} X_{n-i})}]$$

$$\leq \min_{s>0} e^{-s(d+(v-1)/R)} (\mathbb{E}[e^{sX_n}])^{v+1}$$

$$\leq \min_{0<s<\mu} \frac{e^{-s(d+(v-1)/R)}}{(1 - s/\mu)^{v+1}}. \tag{13}$$

In the second step we used Chernoff bound and the fact that $\frac{1}{R} \leq d$. In step three we used that the $\{X_n\}$ are i.i.d. while step four uses the characteristic function of exponential service with rate μ. Using (13) in the summation we obtain

$$\sum_{v=K}^{\infty} \Phi(v,R) \leq \sum_{v=K}^{\infty} \min_{0<s<\mu} \frac{e^{-s(d+(v-1)/R)}}{(1 - s/\mu)^{v+1}}$$

$$\leq \min_{0<s<\mu} \frac{e^{-s(d-1/R)}}{1 - s/\mu} \sum_{v=K}^{\infty} \left(\frac{e^{-s/R}}{1 - s/\mu}\right)^v$$

$$= \min_{0<s<\mu} \frac{e^{-s(d+(K-1)/R)}}{(1 - s/\mu)^K (1 - s/\mu - e^{-s/R})}.$$

In the above step we have used the fact that, for $R < \mu$ there exists an s such that $\frac{e^{-s/R}}{1-s/\mu} < 1$. This can be quickly verified by showing that there exists an s such that logarithm of the fraction, given by $-s/R - \log(1 - s/\mu)$ is negative. To see this, note that for small s, $\log(1 - s/\mu)$ can be accurately approximated by $\frac{-s}{\mu}$. Therefore, for $R < \mu$, we can always find s close to 0 such that $-s/R - \log(1 - s/\mu)$ is negative.

C. Proof of Corollary 1

We show that if K_α satisfies (12), then it satisfies (8).

$$\sum_{v=0}^{\infty} \Phi(v,R) \leq \sum_{v=0}^{K_\alpha-1} \Phi(v,R) + \sum_{v=K_\alpha}^{\infty} \Phi(v,R)$$

$$\leq \sum_{v=0}^{K_\alpha-1} \Phi(v,R) + \min_{0<s<\mu} \frac{e^{-s(d+(K_\alpha-1)/R)}}{(1 - s/\mu)_\alpha^K (1 - s/\mu - e^{-s/R})}$$

$$\leq \alpha \sum_{v=0}^{K_\alpha-1} \Phi(v,R).$$

In the second step above we have used Lemma 3.