

Statistical Guarantee Optimization for AoI in Single-Hop and Two-Hop FCFS Systems with Periodic Arrivals

Jaya Prakash Champati, *Member, IEEE*, Hussein Al-Zubaidy, *Senior Member, IEEE*,
and James Gross, *Senior Member, IEEE*

Abstract—Age of Information (AoI) has proven to be a useful metric in networked systems where timely information updates are of importance. In the literature, minimizing “average age” has received considerable attention. However, various applications pose stricter age requirements on the updates which demand knowledge of the AoI distribution. Furthermore, the analysis of AoI distribution in a multi-hop setting, which is important for the study of Wireless Networked Control Systems (WNCS), has not been addressed before. Toward this end, we study the distribution of AoI in a WNCS with two hops and devise a problem of minimizing the tail of the AoI distribution with respect to the frequency of generating information updates, i.e., the sampling rate of monitoring a process, under first-come-first-serve (FCFS) queuing discipline. We argue that computing an exact expression for the AoI distribution may not always be feasible; therefore, we opt for computing upper bounds on the tail of the AoI distribution. Using these upper bounds, we formulate Upper Bound Minimization Problems (UBMP), namely, Chernoff-UBMP and α -relaxed Upper Bound Minimization Problem (α -UBMP), where $\alpha > 1$ is an approximation factor, and solve them to obtain “good” heuristic rate solutions for minimizing the tail. We demonstrate the efficacy of our approach by solving the proposed UBMPs for three service distributions: geometric, exponential, and Erlang. Simulation results show that the rate solutions obtained are near optimal for minimizing the tail of the AoI distribution for the considered distributions.

Index Terms—Age of Information; tail distribution; deterministic arrivals; rate optimization; stochastic network calculus; multi-hop networking

I. INTRODUCTION

In the recent past, there has been an ever increasing interest in studying Wireless Networked Control Systems (WNCS) that support time-critical-control applications which include, among many others, autonomous vehicle systems, automation of manufacturing processes, smart grids, Internet-of-Things (IoT), sensor networks and augmented reality. A basic building block in WNCS is

J.P. Champati, H. Al-Zubaidy, and J. Gross are with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {jpra,hzubaidy,jamesgr}@kth.se). This research was supported in part by the Swedish Research Council (VR) under grant 2016-04404.

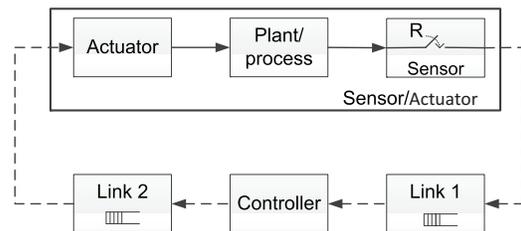


Fig. 1: A networked control system with a remote controller.

depicted in Figure 1. A sensor samples a plant/process of interest and transmits the status updates or packets over a wireless channel (link 1) to a controller. The controller computes a control input using the received status update and transmits it to an actuator, using another communication channel (link 2). A status update that is received at the controller after a certain duration of its generation time may become stale, and the control decision taken based on this stale sample may result in untimely actuation affecting the performance of a time-critical-control application in a WNCS. Similarly, the same effect could result from a control decision (based on a fresh status update) reaching the actuator after a delay deadline. In this respect, the traditional goal of maximizing throughput becomes less relevant as freshness of the status updates not only depends on queuing and transmission delays in the network, but also on the frequency of generating updates at the source.

Age of Information (AoI), proposed in [1], has emerged as a relevant performance metric in quantifying the freshness of the status updates at a destination. It is defined as the time elapsed since the generation of the latest status update received at the destination. AoI accounts for the frequency of generation of updates by the source, since it linearly increases with time until a status update with latest generation time is received at the destination. Whenever such an update is received, AoI resets to the system delay of the update indicating its age.

Motivated by the fact that having access to fresher status updates improves the control performance in WNCS, we model the control network by a two-hop FCFS queuing system and formulate a problem of computing optimal sampling rate that minimizes AoI in this system¹. Several research works in the recent past addressed the problem of optimizing sampling rate in different queuing systems under various settings. However, as we explain in Section II, these works either consider a single-hop system or memoryless arrivals or some form of “average age” function. In contrast, we consider two novel aspects that are relevant to time-critical-control applications. First, we consider periodic arrival process by assuming that the process of interest is sampled at a constant rate R . This is motivated by the fact that sensors in practice are typically configured to generate samples periodically. Second, since optimizing average statistics of AoI may not meet stringent QoS requirements, for instance, in a safety-critical system [3], we consider optimizing *AoI violation probability*, i.e., the probability that AoI at the actuator violates a given *age limit* d . This metric represents, for example, a reliability measure constraint that is required at the actuator to insure that the state of the plant remains within a predetermined safety boundary. Furthermore, in a WNCS, an absolute guarantee (i.e., reliability of 1) may not be possible due to variability of the wireless channel and only probabilistic guarantees can be provided. This motivated us to use the distribution of AoI as a metric rather than other frequently used metrics in the literature, e.g., peak AoI and average AoI.

We consider a heterogeneous network, i.e., server at the first queue and server at the second queue may have different service-time distributions. The queues operate using First-Come-First-Serve (FCFS) scheduling discipline. We note that, in the AoI literature, different scheduling disciplines are considered: for example [4]–[6] considered FCFS, [7]–[9] considered LCFS, and [10], [11] considered packet management schemes such as using a unit capacity queue with packet replacement. Our motivation for considering FCFS discipline in this work is the following. First, analysis and optimization of AoI violation probability under FCFS is an open problem. Second, it is not only an interesting problem from academic (queuing-theoretic) point of view, due to FCFS being more intuitive and hence such analysis being more comprehensible, but also important from practical point of view as most queues in practice operate under FCFS. Third, key insights, e.g., as the sampling rate R increases, in contrast to *delay*, AoI first decreases and then increases [4], that are established under FCFS discipline may be extended to other disciplines as well. Lastly, the

analysis of many important queuing disciplines can be based on, and sometimes directly derived from, that of FCFS, i.e., by introducing a queue reordering stage based on arrival instance, priority, or some fairness parameter before serving the head of the queue. We believe that our analysis can potentially be extended to such queuing disciplines in future works.

Assuming that the processing time at the controller is negligible, we aim to compute R that minimizes the *AoI violation probability* at the egress of the second queue. As we will see in a while, an exact expression for the AoI violation probability in a two-hop network with periodic arrivals and general service-time distributions is intractable. Therefore, we resort to working with tractable upper bounds which facilitate the computation of “good” heuristic solutions. In particular, we first compute the upper bounds for the single-hop case, i.e., the D/G/1 queue, due to its relevance in applications where both controller and actuator are collocated. We formulate the Upper Bound Minimization Problems (UBMP) and use them to compute heuristic rate solutions for AoI violation probability minimization. We then extend the results for two-hop and N-hop tandem queuing systems using max-plus convolution for the service processes.

The main contributions of this work are summarized below:

- We characterize, for the first time, the probability that AoI violates a given age limit d for a single-source single destination multi-hop network under FCFS, and assuming periodic source where packets are generated at a constant rate R .
- 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 tagged packet (defined in Section V) over the rate region $[\frac{1}{d}, \mu)$, where μ is the service capacity of the network.
- Using the above characterization, we first propose a UBMP for the single-hop scenario, i.e., the D/G/1 queue. Noting that the objective function in the UBMP can be intractable, we propose a Chernoff-UBMP, that has a closed-form objective, and an α -relaxed UBMP the solution of which has $\alpha > 1$ approximation ratio (worst-case ratio) with respect to the objective function of the UBMP.
- We extend the derived results and formulations for the two-hop queuing system and N -hop tandem queuing system, and present example computation of the expressions for the case of two-hop for geometric, exponential, and Erlang service-time distributions.
- We demonstrate the efficacy of the heuristic solutions provided by Chernoff-UBMP and α -relaxed UBMP using simulation for different service-time

¹A preliminary version of this work only considering single-hop scenario appeared in [2].

distributions. Finally, we present simulation results comparing the performance of FCFS with queue management policies that use unit buffer and packet replacement.

The rest of the paper is organized as follows. In Section II, we present the related work. In Section III, we present the problem formulation. Analysis on the AoI violation probability is presented in Section IV. The UBMP formulations for single-hop, two-hop, and N-hop scenarios are presented in Sections V and VI, respectively. We present the computation of the upper bounds for different service-time distributions in Section VII. Numerical results are presented in Section VIII and we finally conclude in Section IX.

II. RELATED WORK

Several works in the AoI literature have focused on analyzing and providing expressions for average AoI statistics in different queuing systems, e.g., see [7], [12]–[15]. The authors in [10] studied the M/M/1/1 and M/M/1/2*² systems, and computed the average AoI and the distribution of the peak AoI. In contrast, the authors in [11], [16] provided expressions for the distribution of AoI. However, for the case of periodic arrivals, closed-form expressions are provided only for single-hop scenario and for exponential service times in [16], and for the case of no queue in [11]. Next, we summarize works that consider optimizing AoI under different system settings. An interested reader may also refer to [17] and [18] for a comprehensive survey of recent work in this area.

In [4], the authors have addressed the problem of computing the optimal arrival rate to minimize the *time-average age* for M/M/1, M/D/1 and D/M/1 queuing systems. This problem was addressed for M/M/1 with multiple sources in [19]. Several research works that followed considered different design choices including the arrival rate [5], inter-arrival time distribution for a given arrival rate and/or service-time distribution for a given service rate [15], [20]–[22], under different scheduling disciplines and optimized average AoI or average peak AoI in a single-source-single-server system. In [23], preemptive Last Generated First Served (LGFS) policy was shown to minimize the age process in a multi-server single-hop system with exponential service times. An alternative approach to the above works, the generate-at-will source model, was studied in [24]–[26] where generation of a status update can be completely controlled. While the authors in [24] solved for optimal-waiting times between generation times to minimize the average AoI, the authors in [25] solved the problem for any non-decreasing function of AoI,

²A unit capacity queue that holds the latest update.

and the authors in [26] solved the problem of minimum achievable peak AoI in any single-source-single-server system. The authors in [6] studied average AoI and average peak AoI minimization for multiple source-destination links in a wireless network with interference constraints. They used the method of minimizing upper bounds as a means to show that optimal rate design and optimal link scheduling can be separated and provided performance guarantees for the proposed solutions.

In addition to the above, the following literature considered multi-hop settings. For a line network with a single source and no queues, under Poisson arrivals and exponential service times, expressions were derived in [9] for moments, Moment Generating Function (MGF), and stationary distribution of AoI for preemptive last-come-first-served policy. In [8], optimal queuing policies were investigated for a multi-hop network for any arrival sequence and service-time distributions. It was shown that, among non-preemptive policies, LGFS minimizes age processes, in stochastic ordering sense, at all the nodes. The authors in [27], [28] studied average AoI and average peak AoI minimization in a multi-hop wireless network with interference constraints and with packet flows between multiple source-destination pairs assuming that transmission time of a packet equals a unit time slot. The authors in [29] studied average AoI for L -hop multicast network with a single source, n nodes in the first hop, n^2 nodes in the second hop, and so on n^L nodes in the the last hop, with each node having a shifted exponential service time.

Optimizing AoI was also extensively studied for the systems with energy-harvesting source, e.g., see [24], [30], [31]. In the context of a cloud gaming system the authors in [32] used the D/G/1 system model to study the effect of freshness on video frame rendering to the client. Specifically, they have analyzed the average age by considering the aspect of missing frames. In contrast to all the above works, with motivations from the sensor-controller-actuator system in WNCS we study the problem of AoI violation probability minimization in a two-hop queuing system with periodic arrivals.

III. SYSTEM MODEL AND PROBLEM STATEMENT

Motivated by the sensor-controller-actuator communicating over wireless channels, we study a two-hop queuing system, shown in Figure 2, under FCFS scheduling. The 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 nodes by $k \in \{1, 2\}$, and the packets by $n \in \{0, 1, 2, \dots\}$. Let $A_k(n, R)$ denote the arrival instant of packet n and $D_k(n, R)$ the corresponding departure instant at node k . For notational simplicity, we use $A(n, R) = A_1(n, R)$

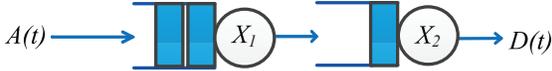


Fig. 2: Model of the two-hop network.

and $D(n, R) = D_2(n, R)$ to denote the arrivals and departures of the system, respectively. Also, we have $A_2(n, R) = D_1(n, R)$. The arrival time of packet n to the system is given by $A(n, R) = \frac{n}{R}$. The service time for packet n at node k is given by a random variable X_k^n . For $k \in \{1, 2\}$, we assume X_k^n are i.i.d., for all n , with mean service rate $\mu_k = \frac{1}{\mathbb{E}[X_k^1]} > 0$. Also, we assume that X_1^n and X_2^n are independent, for all n , but may have non-identical distributions, i.e., the servers could be heterogeneous. We define $\mu \triangleq \min(\mu_1, \mu_2)$. Later, in Section VI-C, we show how the results can be extended to N -hop tandem queuing network.

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

$$\Delta(t, R) \triangleq t - \max_n \{A(n, R) : D(n, R) \leq t\}. \quad (1)$$

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 states of the system. Given the age limit d , we are interested in solving the following problem \mathcal{P} :

$$\min_R \lim_{t \rightarrow \infty} \mathbb{P}(\Delta(t, R) > d).$$

Let $R^*(d)$ denote an optimal rate solution for \mathcal{P} . In the sequel, we refer to $\lim_{t \rightarrow \infty} \mathbb{P}(\Delta(t, R) > d)$ as *AoI violation probability*.

Henceforth, we drop R from the notation when it is obvious from the context, for the sake of notation simplicity. For $k \in \{1, 2\}$, the MGF of X_k^n is given by $M_k(s) = \mathbb{E}[e^{sX_k^n}]$.

We now state the Chernoff bound, which will be used extensively to formulate the upper bound minimization problems in the sequel. Assuming that the moment generating function of a random variable Y exists, the **Chernoff bound** for its distribution is given by

$$\mathbb{P}\{Y > y\} \leq \min_{s>0} e^{-sy} \mathbb{E}[e^{sY}].$$

Note that the upper bounds derived using the Chernoff bound involves minimization over the parameter s . We shall see that, for the two-hop network, these bounds attain finite values only when there exists $s > 0$ such that $\max(M_1(s), M_2(s)) < e^{s/R}$. To this end, we formulate the minimization problems over the set $\mathcal{S} \subseteq \mathbb{R}^+$ which

characterizes s values for which $\max(M_1(s), M_2(s)) < e^{s/R}$, i.e.,

$$\mathcal{S} \triangleq \{s > 0 : \max(M_1(s), M_2(s)) < e^{s/R}\}. \quad (2)$$

We assume that \mathcal{S} is non-empty. In the following lemma we show that this assumption is in fact a sufficient condition for the stability of the system.

Lemma 1. *If there exists $s > 0$ such that*

$$\max(M_1(s), M_2(s)) < e^{s/R},$$

then the queues are stable.

Proof. Recall that the queues are stable if $\min(\mu_1, \mu_2) > R$. Consider the case $M_1(s) < e^{s/R}$, which implies

$$\mathbb{E}[e^{sX_1^n}] < e^{s/R} \Rightarrow e^{s\mathbb{E}[X_1^1]} < e^{s/R} \Rightarrow \mu_1 > R,$$

for any $s > 0$. In the second step above we have used Jensen's inequality. Similarly, if $M_2(s) < e^{s/R}$, then $\mu_2 > R$. Therefore, if there exists $s > 0$ such that $\max(M_1(s), M_2(s)) < e^{s/R}$, then $\min(\mu_1, \mu_2) > R$, and the lemma follows. \square

We define

$$\beta_k(s) \triangleq \frac{M_k(s)}{e^{s/R}}, \quad k \in \{1, 2\}. \quad (3)$$

By definition, for all $s \in \mathcal{S}$, $\beta_k(s) < 1$. The list of symbols used in the paper are summarized in Table I.

TABLE I: List of Symbols

k	Node/link index
N	Number of nodes
n	Packet index
R	Sampling rate
T	Inter-arrival time ($\frac{1}{R}$)
$A_k(n, R)$	Arrival time of packet n at node k
$D_k(n, R)$	Departure time of packet n at node k
$A(n, R)$	Arrival time of packet n in the system ($\frac{n}{R}$)
$D(n, R)$	Departure time of packet n from the system
X_k^n	Service time of packet n at node k
μ_k	Service rate at node k
$\Delta(t, R)$	Age of information at time t
d	Age limit
$M_k(\cdot)$	MGF of service at node k
\hat{n}_R	Index of the first arrival after time $t - d$

IV. AOI VIOLATION PROBABILITY ANALYSIS

In this section, we study the properties of the distribution of AoI – the results derived are valid for any number of nodes in tandem between the source and the destination given that the packets are input to the network by the source at a constant rate R and the network uses FCFS.

We start by investigating structural characteristics of the stochastic behaviour of AoI. Toward this end, we use the max-plus representation of Reich's equation to model the evolution of the queues. For any realization of the service times at node k , the relation between $D_k(n, R)$, $A_k(n, R)$ and $\{X_k^n\}$, is given by [33]:

$$D_k(n, R) = \max_{0 \leq v \leq n} \{A_k(n-v, R) + \sum_{i=0}^v X_k^{n-i}\}. \quad (4)$$

We note that equation (4) is a direct consequence of using recursion on a fundamental relation in queuing system: $D_k(n, R) = \max\{D_k(n-1, R), A_k(n)\} + X_k^n$, which states that the departure time of packet n is given by either its service time plus departure time of previous packet $n-1$ or arrival time of packet n plus its service time, whichever is greater.

Consider the definition in (1), for $\Delta(t, R)$ not to exceed the age limit d , the latest departure at t must have arrived no earlier than $t-d$. Therefore, to study the distribution of $\Delta(t, R)$, 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 index of the first arrival since time $t-d$, given by

$$\hat{n}_R \triangleq \lceil R(t-d) \rceil. \quad (5)$$

The tagged packet³ \hat{n}_R plays a key role in characterizing the violation probability as we will show next.

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

Lemma 2. *Given the input arrival rate R , age limit d , and $t < \infty$, if there exists n such that $t-d \leq \frac{n}{R} < t$, then $\mathbb{P}\{\Delta(t, R) > d\} = \mathbb{P}\{D(\hat{n}_R) > t\}$, otherwise, $\mathbb{P}\{\Delta(t, R) > d\} = 1$.*

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

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., $A(n_R^*) < t-d$. Therefore,

$$\mathbb{P}(\Delta(t, R) > d) = \mathbb{P}(t - A(n_R^*) > d) = 1.$$

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 3. In this case, we show that the event $\{\Delta(t, R) \leq d\}$ is equivalent to the event $\{D(\hat{n}_R) \leq t\}$. Suppose that the event $\{\Delta(t, R) \leq d\}$ occurred, then $A(n_R^*) \geq t-d$. By definition of \hat{n}_R , we

³ \hat{n}_R is a function of $t-d$ as well. We omit $t-d$ from the notation here for ease of exposition.

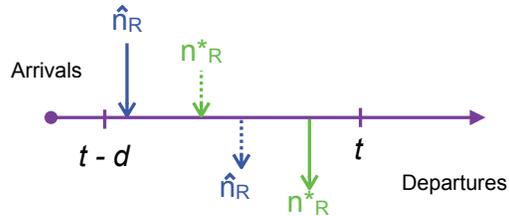


Fig. 3: Time-line of events for Case 2 in Lemma 2 proof.

should have $A(\hat{n}_R) \leq A(n_R^*)$ which implies $D(\hat{n}_R) \leq D(n_R^*) \leq t$, due to FCFS assumption. Therefore,

$$\{\Delta(t, R) \leq d\} \subseteq \{D(\hat{n}_R) \leq t\}. \quad (6)$$

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

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

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

The intuition behind the result in Lemma 2 is that AoI exceeds d at time t if none of the packets generated/arrived in the time interval $[t-d, t]$ have reached the destination at time t or no packets are generated in this interval. Note that **Case 1** in the above proof essentially represents an under-sampling of the process under observation, i.e., at the current time t the sampling rate R is simply too low such that there is no packet generated in the time interval $[t-d, t]$.

We next present the steady-state results for the two-hop system based on the result obtained in Lemma 2.

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

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

$$\lim_{t \rightarrow \infty} \mathbb{P}\{\Delta(t, R) > d\} = \lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\}. \quad (7)$$

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

$$\limsup_{t \rightarrow \infty} \mathbb{P}\{\Delta(t, R) > d\} = 1,$$

$$\liminf_{t \rightarrow \infty} \mathbb{P}\{\Delta(t, R) > d\} = \lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\}.$$

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

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

Furthermore, since $t-d \leq A(\hat{n}_R)$ 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 2 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 . To see this, for a given n consider the two time intervals $(\frac{n}{R}, \frac{n}{R} + d]$ and $(\frac{n}{R} + d, \frac{n+1}{R})$. Note that the latter time interval is non-empty because $d < \frac{1}{R}$. Now, for time instants $t \in (\frac{n}{R}, \frac{n}{R} + d]$ we have $t-d \leq \frac{n}{R} < t$, and therefore for such t using Lemma 2 we have $\mathbb{P}\{\Delta(t, R) > d\} = \mathbb{P}\{D(\hat{n}_R) > t\}$. On the other hand, for time instants $t \in (\frac{n}{R} + d, \frac{n+1}{R})$, there is no n value such that $t-d \leq \frac{n}{R} < t$ is true, and therefore from Lemma 2 we have $\mathbb{P}\{\Delta(t, R) > d\} = 1$. This implies that as t goes to infinity the violation probability either equals $\mathbb{P}\{D(\hat{n}_R) > t\}$ or 1 depending on the value of t . Thus, we obtain the limit supremum and the limit infimum. \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 $\mu = \min(\mu_1, \mu_2)$, and $R < \mu$ ensures queue stability. In light of this, and using (7) from Theorem 1, we formulate an equivalent problem $\tilde{\mathcal{P}}$ as follows:

$$\min_{\frac{1}{d} \leq R < \mu} \lim_{t \rightarrow \infty} \mathbb{P}(D(\hat{n}_R) > t). \quad (8)$$

Remark 1: The results in Lemma 2 and Theorem 1 are valid for arbitrary single-source single-destination network with constant arrival rate R and using FCFS queuing discipline, i.e., packets should be received at the destination in the same order as they are transmitted by the source. For arbitrary network topology, one can formulate problem $\tilde{\mathcal{P}}$ given in (8) with the following constraints on R : 1) $R \geq \frac{1}{d}$, and 2) R belongs to the rate region in which the network is stable.

Next, we present our solution approach for solving $\tilde{\mathcal{P}}$ for a single-hop case and then show how the approach can be extended for the two-hop system in Section VI.

V. SINGLE-HOP SCENARIO

In this section we solve $\tilde{\mathcal{P}}$ by assuming that $X_2^n = 0$ for all n . This implies that $D(n) = D_1(n)$, $\mu_2 = \infty$

and the system is equivalent to the D/GI/1 system. Our motivation for presenting the single-hop case is because of its importance in solving the two-hop case, and also due to its relevance to practical scenarios, where only estimation of the processes is required, or both controller and actuator are collocated. In order to find a solution for $\tilde{\mathcal{P}}$, we must first evaluate the probability $\mathbb{P}\{D(\hat{n}_R) > t\}$, where $D(n)$ is given by (4). Note that $D(n)$ is random, since the service process $\{X_1^n, n \geq 0\}$ 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 $D(\hat{n}_R)$. Consequently, we propose the Upper Bound Minimization Problem (UBMP) and its more computationally tractable counterparts α -UBMP and Chernoff-UBMP to obtain near optimal heuristic solutions for $\tilde{\mathcal{P}}$.

A. A Bound for the Distribution of D

As mentioned earlier, the evaluation of the distribution function of $D(n)$ 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., $\mathbb{P}\{\max_i Y_i > y\} \approx \max \mathbb{P}\{Y_i > y\}$. 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 also provides good heuristic solutions for $\tilde{\mathcal{P}}$. In the following lemma, we present this upper bound for the distribution function $\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\}$.

Lemma 3. Given d , we have

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

where

$$\Phi(v, R) \triangleq \mathbb{P}\left\{\sum_{i=0}^v X_1^i > d + \frac{v-1}{R}\right\}. \quad (9)$$

Proof. Using (4), we have

$$\mathbb{P}\{D(\hat{n}_R) > t\} = \mathbb{P}\left\{\max_{0 \leq v \leq \hat{n}_R} \left(A(\hat{n}_R - v) + \sum_{i=0}^v X_1^{\hat{n}_R - i}\right) > t\right\}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \sum_{v=0}^{\hat{n}_R} \mathbb{P} \left\{ \sum_{i=0}^v X_1^{\hat{n}_R-i} > t - \frac{\hat{n}_R - v}{R} \right\} \\
&\stackrel{(b)}{\leq} \sum_{v=0}^{\hat{n}_R} \mathbb{P} \left\{ \sum_{i=0}^v X_1^{\hat{n}_R-i} > t - \frac{R(t-d) + 1 - v}{R} \right\} \\
&= \sum_{v=0}^{\hat{n}_R} \underbrace{\mathbb{P} \left\{ \sum_{i=0}^v X_1^i > d + \frac{v-1}{R} \right\}}_{\triangleq \Phi(v, R)}.
\end{aligned}$$

In step (a) we have applied the union bound, and used $\hat{n}_R = \lceil R(t-d) \rceil \leq R(t-d) + 1$ in step (b). The result follows by noting that \hat{n}_R goes to infinity as t goes to infinity. \square

B. UBMP Formulations

Using (8), Lemma 3 and $\mu_2 = \infty$, we obtain the following UBMP problem.

$$\min_{\frac{1}{d} \leq R < \mu_1} \sum_{v=0}^{\infty} \Phi(v, R). \quad (10)$$

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 .

A shortcoming of UBMP is that its objective function is intractable, in general, as it involves computation of a sum of infinite terms and each term requires computation of the distribution of sum of service times. To this end, we formulate Chernoff-UBMP obtained by using Chernoff bound for $\Phi(v, R)$ in Lemma 3.

1) *Chernoff-UBMP*: Since X_1^n are i.i.d, the Chernoff bound for $\Phi(v, R)$, defined in (9), is given by

$$\Phi(v, R) \leq \min_{s \in \mathcal{S}} e^{-s(d + \frac{v-1}{R})} \mathbb{E}[e^{s \sum_{i=0}^v X_1^i}] \quad (11)$$

$$\begin{aligned}
&= \min_{s \in \mathcal{S}} e^{-s(d + \frac{v-1}{R})} M_1^{v+1}(s) \\
&= \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} M_1(s) \beta_1^v(s), \quad (12)
\end{aligned}$$

where $\beta_1(s) = \frac{M_1(s)}{e^{s/R}}$ (defined in (3)), and \mathcal{S} is defined in (2). Recall that, $\beta_1(s) < 1$ for all $s \in \mathcal{S}$. Therefore, using (11) in the result of Lemma 3, we obtain

$$\sum_{v=0}^{\infty} \Phi(v, R) \leq \sum_{v=0}^{\infty} \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} M_1(s) \beta_1^v(s) \quad (13)$$

$$\begin{aligned}
&\leq \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} M_1(s) \sum_{v=0}^{\infty} \beta_1^v(s) \\
&= \min_{s \in \mathcal{S}} \underbrace{e^{-s(d - \frac{1}{R})} \cdot \frac{M_1(s)}{(1 - \beta_1(s))}}_{\triangleq \Psi_1(s, d, R)}. \quad (14)
\end{aligned}$$

Even though the Chernoff bound relaxes the upper bound in Lemma 3, its objective function has a closed-form expression and can be computed numerically. The

following theorem immediately follows from (13) and Lemma 3.

Theorem 2. Given d , an upper bound for the violation probability for a single hop is given by

$$\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} \leq \min_{s \in \mathcal{S}} \Psi_1(s, d, R),$$

where $\Psi_1(s, d, R)$ is defined in (13).

With a slight abuse in the usage, we refer to the bound given in Theorem 2 as *Chernoff bound*. In the following, we formulate the Chernoff-UBMP for the single-hop scenario:

$$\min_{\frac{1}{d} \leq R < \mu_1} \min_{s \in \mathcal{S}} \Psi_1(s, d, R). \quad (15)$$

Lemma 4. The function $\Psi_1(s, d, R)$ is strictly convex with respect to $\frac{1}{R}$.

Proof. Recall that $T = \frac{1}{R}$. We prove that $\frac{\partial^2 \Psi_1(s, d, T)}{\partial T^2} > 0$ for all $s \in \mathcal{S}$. Let us define $f(T)$ as follows.

$$f(T) = \frac{e^{2sT}}{(e^{sT} - M_1(s))}.$$

Then, we rewrite $\Psi_1(s, d, T)$ as follows.

$$\Psi_1(s, d, T) = e^{-sd} [M_1(s)] \cdot f(T).$$

From the above equation we infer that it is sufficient to prove $\frac{\partial^2 f(T)}{\partial T^2} > 0$. Taking first derivative $f'(T) = \frac{\partial f(T)}{\partial T}$, we obtain

$$\begin{aligned}
f'(T) &= \frac{2se^{2sT}}{(e^{sT} - M_1(s))} - \frac{e^{2sT} 2se^{sT}}{(e^{sT} - M_1(s))^2} \\
&= sf(T) \left[1 - \frac{M_1(s)}{e^{sT} - M_1(s)} \right]. \quad (16)
\end{aligned}$$

Taking the second derivative $f''(T) = \frac{\partial^2 f(T)}{\partial T^2}$, we obtain

$$\begin{aligned}
f''(T) &= sf'(T) \left[1 - \frac{M_1(s)}{e^{sT} - M_1(s)} \right] + \frac{s^2 f(T) M_1(s) e^{sT}}{(e^{sT} - M_1(s))^2} \\
&= s^2 f(T) \left[1 - \frac{M_1(s)}{e^{sT} - M_1(s)} \right]^2 + \frac{s^2 f(T) M_1(s) e^{sT}}{(e^{sT} - M_1(s))^2} > 0.
\end{aligned}$$

In the second step above we have used (16). The last step follows by noting that $e^{sT} > M_1(s)$ for all $s \in \mathcal{S}$, $M_1(s) > 0$ for all s , and $f(T) > 0$. \square

Lemma 5. For $s > 0$, the function $\Psi_1(s, d, R)$ is convex in s .

Proof. We have

$$\begin{aligned}
\Psi_1(s, d, R) &= e^{-s(d - \frac{1}{R})} \cdot \frac{M_1(s)}{(1 - \beta_1(s))} \\
&= e^{-s(d - \frac{1}{R})} \cdot \sum_{v=0}^{\infty} M_1(s) \beta_1^v(s) \\
&= \sum_{v=0}^{\infty} e^{-s(d + \frac{v-1}{R})} M_1^{v+1}(s) = \sum_{v=0}^{\infty} (\mathbb{E}[e^{-s\hat{X}}])^{v+1},
\end{aligned}$$

where $\hat{X} = (d + \frac{v-1}{R})/(v+1) - X_1^1$. Recall that the sum of convex functions is a convex function. Therefore, from the last step above, we infer that $\Psi_1(s, d, R)$ is convex if $(\mathbb{E}[e^{-sX}])^{v+1}$ is convex for $v \in \{0, 1, \dots\}$. For $s > 0$, $e^{-s\hat{X}}$ is convex in s for any v and any realization of X_1^1 . Therefore, $\mathbb{E}[e^{-s\hat{X}}]$ is convex, and since x^{v+1} is convex and increasing in x , we have that $(\mathbb{E}[e^{-sX}])^{v+1}$ is convex. Hence the result is proven. \square

Both Lemmas 4 and 5 can be leveraged to efficiently solve (15). The heuristic solutions we obtain by solving the Chernoff-UBMP can be improved further for service distributions for which the distribution of a finite sum of service times can be computed exactly. Therefore, we next propose a relatively tight upper bound called α -relaxed upper bound and formulate α -UBMP.

2) α -UBMP: In the upper bound provided in Lemma 3, we propose to compute first $K < \infty$ terms of the summation, and use Chernoff bound for the rest of the terms. In the following, we make this precise. We first present a bound on the summation starting from K .

Lemma 6. *For any $K \geq 0$, we have*

$$\sum_{v=K}^{\infty} \Phi(v, R) \leq \min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s).$$

Proof. The result follows by using the upper bound for $\Phi(v, R)$ given in (11) and repeating the steps in (13) for the summation over v from K to infinity. \square

For the single hop scenario we define α as follows.

$$\alpha = 1 + \frac{\min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s)}{\sum_{v=0}^{K-1} \Phi(v, R)}.$$

Note that α depends on the value of K . Using Lemmas 3 and 6, we next state the α -relaxed upper bound without proof.

Theorem 3. *Given d , the α -relaxed upper bound for the violation probability for a single hop is given by*

$$\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} \leq \sum_{v=0}^{K-1} \Phi(v, R) + \min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s).$$

Note that, by definition the α -relaxed upper bound is at most α times worse than the upper bound $\sum_{v=0}^{\infty} \Phi(v, R)$. More precisely, the α -relaxed upper bound has α approximation factor with respect to $\sum_{v=0}^{\infty} \Phi(v, R)$. To see this,

$$\begin{aligned} & \sum_{v=0}^{K-1} \Phi(v, R) + \min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s) \\ &= \sum_{v=0}^{K-1} \Phi(v, R) \left(1 + \frac{\min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s)}{\sum_{v=0}^{K-1} \Phi(v, R)} \right) \end{aligned}$$

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

Note that $\alpha > 1$, and it is easy to see that as K increases, the value of α approaches 1 from above. In this work, we choose K the largest value that is computationally tractable in numerical evaluations. Now, we formulate α -UBMP as follows:

$$\min_{\frac{1}{d} \leq R < \mu_1} \sum_{v=0}^{K-1} \Phi(v, R) + \min_{s \in \mathcal{S}} \Psi_1(s, d, R) \beta_1^K(s). \quad (17)$$

VI. EXTENSIONS TO TWO-HOP AND N-HOP SCENARIOS

In this section, we present Chernoff-UBMP and α -UBMP for the two-hop scenario and also present Chernoff-UBMP for N-hop tandem queuing network.

In the following we first focus on the two-hop scenario. Similar to the case of single-hop scenario, we use Reich's equation and apply union bound to obtain an upper bound for the AoI violation probability which is presented in the following lemma.

Lemma 7. *Given d , and \hat{n}_R as defined in (5), we have*

$$\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} \leq \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R-v_0} \Phi(v_0, v_1, R),$$

where

$$\Phi(v_0, v_1, R) \triangleq \mathbb{P}\left\{ \sum_{i=0}^{v_0} X_2^i + \sum_{i=0}^{v_1} X_1^i > d + \frac{v_0 + v_1 - 1}{R} \right\}. \quad (18)$$

Proof. The proof is given in Appendix A. \square

A. Chernoff-UBMP for Two-Hop Scenario

Theorem 4. *For the two-hop network with deterministic arrivals, the violation probability is upper bounded as follows:*

$$\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} \leq \min_{s \in \mathcal{S}} \Psi_2(s, d, R),$$

where

$$\Psi_2(s, d, R) = \frac{e^{-s(d - \frac{1}{R})} M_1(s) M_2(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}. \quad (19)$$

Proof. We use the relation between departure times, arrival times and the service times given by (4) iteratively and apply union bound and Chernoff bound to obtain the result. The details of the proof are given in Appendix B. \square

The Chernoff-UBMP problem for the two-hop network is stated below:

$$\min_{\frac{1}{d} \leq R < \mu} \min_{s \in \mathcal{S}} \Psi_2(s, d, R). \quad (20)$$

The lemmas below provide convexity properties of $\Psi_2(s, d, R)$. Since the proofs of the lemmas are similar

to that in the case of single-hop scenario (Lemmas 4 and 5), we omit them here.

Lemma 8. For the two-hop network with deterministic arrivals, given $s \in \mathcal{S}$ and $d > 0$, $\Psi_2(s, d, R)$ is convex with respect to $\frac{1}{R}$.

Lemma 9. For the two-hop network with deterministic arrivals, given $s \in \mathcal{S}$ and $d > 0$, $\Psi_2(s, d, R)$ is convex with respect to s .

B. α -UBMP for Two-Hop Scenario

In the following theorem we present the α -relaxed upper bound.

Theorem 5. For the two-hop network with deterministic arrivals, for any $K \geq 1$, the α -relaxed upper bounded is given by

$$\sum_{v_0=0}^{K-1} \sum_{v_1=0}^{K-1} \Phi(v_0, v_1, R) + \min_{s \in \mathcal{S}} \Psi(s, d, R, K).$$

where

$$\Psi(s, d, R, K) \quad (21)$$

$$= e^{-s(d-\frac{1}{R})} M_1(s) M_2(s) \frac{(\beta_1^K(s) + \beta_2^K(s) - \beta_1^K(s)\beta_2^K(s))}{(1-\beta_1(s))(1-\beta_2(s))}. \quad (22)$$

Proof. The proof is given in Appendix C. \square

We note that the α -relaxed upper bound is computationally expensive when compared to that in the single-hop scenario because of the nested sum.

C. N -hop Scenario

For an N -hop tandem network we have $k \in \{1, 2, \dots, N\}$ and $D(n) = D_N(n)$. For simplicity of presentation, in this section, we assume that X_k^n are identically distributed. Therefore, we have $\mu = \mu_k$ for all k , and $M_k(s) = M_1(s)$ for all k . We now define the set \mathcal{S} as follows.

$$\mathcal{S} = \{s > 0 : M_1(s) < e^{s/R}\}.$$

Lemma 10. Given d , and \hat{n}_R as defined in (5), we have

$$\lim_{t \rightarrow \infty} \mathbb{P}(D(\hat{n}_R) > t) \leq \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R-v_0} \dots \sum_{v_{N-1}=0}^{\hat{n}_R-v_{N-2}} \Phi(v_0^{N-1}, R),$$

where

$$\Phi(v_0^{N-1}, R) \triangleq \mathbb{P} \left\{ \sum_{k=0}^{N-1} \sum_{i=0}^{v_k} X_{N-k}^i > d + \frac{\sum_{k=0}^{N-1} v_k - 1}{R} \right\}, \quad (23)$$

and $v_0^{N-1} = (v_0, v_1, \dots, v_{N-1})$.

Proof. The proof follows similar steps as the proof of Lemma 7 and is omitted. \square

Theorem 6. For the N -hop network with deterministic arrivals, the violation probability is upper bounded as follows:

$$\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} \leq \min_{s \in \mathcal{S}} \Psi_N(s, d, R),$$

where

$$\Psi_N(s, d, R) = e^{-s(d-\frac{1}{R})} \cdot \frac{[M_1(s)]^N}{[1 - \frac{M_1(s)}{e^{s/R}}]^N}. \quad (24)$$

Proof. We use the relation between departure times, arrival times and the service times given by (4) recursively starting from the last node N , and apply union bound and Chernoff bound to obtain the result. The proof follows similar steps as in the proof of Theorem 4 and therefore it is omitted. \square

Therefore, an upper bound minimization problem for the N -hop network can be stated as follows:

$$\min_{\frac{1}{d} \leq R < \mu} \min_{s \in \mathcal{S}} \Psi_N(s, d, R). \quad (25)$$

Discussion: We note that similar to the single-hop and two-hop scenario $\Psi_N(s, d, R)$ is also convex with respect to $\frac{1}{R}$ and with respect to s . One may also obtain α -UBMP for the N -hop scenario. However, the α -relaxed upper bound involves the nested sum which becomes computationally expensive as N increases. Furthermore, we note that as N increases the upper bounds become more relaxed and therefore the heuristic solutions provided by Chernoff-UBMP may not be close to optimal solution. Nevertheless, these heuristic solutions could potentially be used as starting points. For example, when the controller has non-negligible processing time, the sensor-controller-actuator can be modelled as a three-hop tandem queuing system and one may use the heuristic solutions provided by the Chernoff-UBMP for three-hop scenario.

Next, we present an independent result for service-time distributions with bounded support.

Service Distributions with Bounded Support: Note that in practice, the service time distributions typically have bounded support. For example, the channel capacity for transmissions is always upper bounded due to bandwidth limitation. Considering that the service time is upper bounded by $b \in \mathbb{R}_{>0}$, in the following theorem we present a result for computing an optimal rate for age limits above certain threshold.

Corollary 1. For an N -hop network, if the support of the service time distribution is upper bounded by $b < \infty$, then for all $d \geq (N+1)b$, the AoI violation probability is zero at $R \leq (N+1)/d$, i.e., these rate solutions are optimal for (8).

Proof. We rewrite $\Phi(v_0^{N-1}, R)$ (defined in (23)) as follows:

$$\Phi(v_0^{N-1}, R) = \mathbb{P} \left\{ \sum_{k=0}^{N-1} \sum_{i=0}^{v_k} \left(X_{N-k}^i - \frac{1}{R} \right) > d - \frac{N+1}{R} \right\}.$$

For $R \leq (N+1)/d$, we have $X_k^n \leq b \leq \frac{1}{R^*}$ for all $k \geq 1$ and for all n , and we obtain

$$\Phi(v_0^{N-1}, R^*) = \mathbb{P} \left\{ \sum_{k=0}^{N-1} \sum_{i=0}^{v_k} \left(X_{N-k}^i - \frac{1}{R} \right) > 0 \right\} = 0.$$

Therefore, from Lemma 10 we conclude that the AoI violation probability $\lim_{t \rightarrow \infty} \mathbb{P}(T_D(\hat{n}_R) > t)$ is equal to zero when $R \leq (N+1)/d$. \square

VII. APPLICATION EXAMPLES: GEOMETRIC, EXPONENTIAL AND ERLANG SERVICE

In the following we show the computation of the upper bounds for typical service distributions, namely, geometric, exponential and Erlang. These distributions are most commonly used in the queuing analysis, and also they serve as good models for several practical service-time processes. Note that for these distributions, the distribution of the sum of service times is known and thus the α -relaxed upper bound can be computed. Later in Section VIII we will evaluate the performance of the the computed heuristic solutions for these service distributions. To shorten the expressions, in the sequel we denote

$$Y_1 = \sum_{i=0}^{v_1} X_1^i, \quad Y_2 = \sum_{i=0}^{v_0} X_2^i, \quad \text{and} \quad \kappa = d + \frac{v_0 + v_1 - 1}{R}.$$

A. Geometric Service: Wireless Links with Packet Errors

Consider that each packet generated by the sensor is of fixed length and the packets that carry actuator commands are also of fixed length, possibly different from sensor packet length. To accommodate for packet transmission errors in the wireless links, we use geometric distribution to model the number of time slots required for transmitting a packet successfully. In particular, we consider the service distributions at link 1 and link 2 to be geometric with success probabilities p_1 and p_2 , respectively. Given an age limit d at the actuator, we compute R heuristically.

In the following we compute the first term of the α -relaxed upper bound given in Theorem 5. Since Y_1 and Y_2 are integers, we have

$$\begin{aligned} \sum_{v_0=0}^{K-1} \sum_{v_1=0}^{K-1} \Phi(v_0, v_1, R) &= \sum_{v_0=0}^{K-1} \sum_{v_1=0}^{K-1} \mathbb{P}\{Y_1 + Y_2 > \kappa\} \\ &= \sum_{v_0=0}^{K-1} \sum_{v_1=0}^{K-1} \mathbb{P}\{Y_1 + Y_2 > \lfloor \kappa \rfloor\}. \end{aligned} \quad (26)$$

Since for geometrical distribution $X_k^i \geq 1$, for all i and $k \in \{1, 2\}$, we have $Y_1 \geq v_1 + 1$ and $Y_2 \geq v_0 + 1$. Therefore, for $\lfloor \kappa \rfloor \leq v_1 + v_2 + 1$, we have $\mathbb{P}\{Y_1 + Y_2 > \lfloor \kappa \rfloor\} = 1$. For $\lfloor \kappa \rfloor > v_1 + v_2 + 2$ we compute the probability by conditioning on $Y_2 = y$ for positive integers $y \geq v_0 + 1$.

$$\begin{aligned} &\mathbb{P}\{Y_1 + Y_2 > \lfloor \kappa \rfloor\} \\ &= \sum_{y=v_0+1}^{\infty} \mathbb{P}\{Y_1 + Y_2 > \lfloor \kappa \rfloor | Y_2 = y\} \mathbb{P}\{Y_2 = y\} \\ &\quad \mathbb{P}\{Y_1 > \lfloor \kappa \rfloor - y\} \\ &= \sum_{y=v_0+1}^{\lfloor \kappa \rfloor - v_1 - 1} \mathbb{P}\{Y_1 > \lfloor \kappa \rfloor - y\} \mathbb{P}\{Y_2 = y\} + \mathbb{P}\{Y_2 \geq \lfloor \kappa \rfloor - v_1\} \end{aligned}$$

In the last step above we have used $\mathbb{P}\{Y_1 > \lfloor \kappa \rfloor - y\} = 1$ for $y \geq \lfloor \kappa \rfloor - v_1$. Noting that the sum of i.i.d. geometric random variables has a negative binomial distribution, we have

$$\begin{aligned} \mathbb{P}\{Y_2 = y\} &= \mathbb{P}\left\{ \sum_{i=0}^{v_0} X_2^i = y \right\} \\ &= \binom{y-1}{v_0} p_2^{v_0+1} (1-p_2)^{y-v_0-1}, \end{aligned}$$

and

$$\mathbb{P}\{Y_1 > \lfloor \kappa \rfloor - y\} = \frac{B(1-p_2; \lfloor \kappa \rfloor - y - v_1, v_1 + 1)}{B(\lfloor \kappa \rfloor - y - v_1, v_1 + 1)},$$

where $B(\cdot)$ is the incomplete beta function given by

$$\begin{aligned} B(z; a, b) &= \int_0^z x^a (1-x)^b dx, \\ B(a, b) &= \int_0^1 x^a (1-x)^b dx. \end{aligned}$$

Similarly, we compute $\mathbb{P}\{Y_2 \geq \lfloor \kappa \rfloor - v_1\}$. Finally, using $\mathbb{P}\{Y_1 + Y_2 > \lfloor \kappa \rfloor\}$ we compute (26). For computing the Chernoff bound we require the moment generating function, which for geometric service is given below.

$$M_k(s) = \frac{p_k e^s}{1 - (1-p_k)e^s}.$$

Since the Chernoff bound is convex in s , using bisection algorithm we compute its minimum value.

B. Exponential Service

In this subsection, we study the two-hop system with exponentially distributed service times with rates μ_1 and μ_2 at links 1 and 2, respectively. For this case, Y_1 is a sum of $v_1 + 1$ i.i.d. exponential random variables, which is given by the Gamma distribution with shape parameter $v_1 + 1$ and rate parameter μ_1 . Similarly, Y_2 has Gamma distribution with shape parameter $v_2 + 1$ and

rate parameter μ_2 . Therefore, we compute $\Phi(v_0, v_1, R)$ as follows.

$$\begin{aligned}\Phi(v_0, v_1, R) &= \int_0^\infty \mathbb{P}\{Y_1 > \kappa - y\} f_{Y_2}(y) dy \\ &= \int_0^\kappa \mathbb{P}\{Y_1 > \kappa - y\} f_{Y_2}(y) dy + \int_\kappa^\infty f_{Y_2}(y) dy,\end{aligned}\quad (27)$$

where $f_{Y_2}(\cdot)$ is the PDF of Y_2 , given by

$$\begin{aligned}f_{Y_2}(y) &= \frac{\mu_2^{v_2+1} y^{v_2} e^{-\mu_2 y}}{v_2!}, \\ \mathbb{P}\{Y_1 > \kappa - y\} &= \frac{\Gamma(v_1 + 1, \mu_1(\kappa - y))}{v_1!},\end{aligned}$$

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

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

Further, if $\mu_1 = \mu_2 = \mu$, then

$$\Phi(v_0, v_1, R) \Big|_{\mu_1=\mu_2} = \frac{\Gamma(v_0 + v_1 + 2, \mu\kappa)}{(v_0 + v_1 + 1)!}.$$

For computing the Chernoff bound we use the MGF of the exponential distribution which is given below.

$$M_k(s) = \frac{\mu_k}{\mu_k - s}, \text{ for } s < \mu_k.$$

C. Erlang Service

Consider the Erlang service distribution at link k has shape parameter b_k and rate λ_k . This implies $\mu_k = b_k \lambda_k$. We note that, in this case, Y_k has Gamma distribution with shape parameter $(v_k + 1)b_k$ and rate parameter λ_k . Therefore, we compute the bounds using similar expressions given in the previous subsection.

Remark 2: We note that the Chernoff upper bound and the α -relaxed upper bound presented above may take values greater than 1. It is natural to cap the values of these upper bounds by 1 because for probability values an upper bound greater than 1 is not of any use, in general. However, somewhat to our surprise, in our simulations we found that allowing the values of the proposed bounds greater than 1 provides good heuristic solutions for the sampling rate, especially for parameter setting where the upper bounds are always greater than 1. Since our primary objective is to find upper bounds that can provide good heuristic solutions, but need not necessarily be tight upper bounds, we consider values greater than 1 for the bounds in our numerical evaluation. However, this should not be confused with the violation probability which does not exceed 1 at all times.

VIII. NUMERICAL EVALUATION

In this section, we evaluate the performance of α -UBMP solutions and Chernoff-UBMP solutions for geometric, exponential and Erlang service distributions. We first study the trends of the proposed upper bounds in comparison to the AoI violation probability obtained using simulation for both single-hop and two-hop scenarios. We then evaluate the quality of numerically computed solutions using the UBMPs in comparison with that of the simulation-based estimate of the optimum violation probability. Finally, we present simulation results comparing the performance of FCFS with queue management policies that use unit buffer and packet replacement.

Since Chernoff-UBMP is a convex optimization problem we used bisection search, and for α -UBMP we used brute-force search to compute the respective optimal rates. The numerical computations are done using MATLAB, and the simulation is implemented in C where we run 10^{10} iterations for each data point. The default parameters are as follows. For exponential distribution μ_1 and μ_2 equal 1 packet/ms; for Erlang distribution we use shape parameters $b_1 = b_2 = 3$ and rate parameters $\lambda_1 = \lambda_2 = 3$, and therefore the mean rates μ_1 and μ_2 equal one packet/ms; for geometric service we choose success probabilities $p_1 = 0.85$ and $p_2 = 0.9$, and the slot duration is 1 ms. The minimum value for R is chosen to be 0.2 packets/ms and its maximum value is chosen to be $0.75 * \min(\mu_1, \mu_2)$ packets/ms. For all the figures with varying rate R on the x-axis, a constant resolution of 0.025 is used. The minimum value for d is chosen to be 5 ms and its maximum value is chosen to be 15 ms. We use $K = 30$ for computing α -relaxed upper bound for all the distributions because for Geometric service MATLAB does not provide precision guarantees for higher K values for computing $\Phi(v_0, v_1, R)$, and for other service distributions, choosing $K = 30$ is sufficient to obtain α values close to 1.

A. Properties of Upper Bounds

1) *Single Hop:* In Figures 4 and 5, we present the upper bounds and the simulated AoI violation probability for varying arrival rate R and varying age limit d for different distributions for the single-hop scenario. From Figure 4, we observe that the upper bounds and the violation probability have convex nature and a global minimum (highlighted in black circles) in the chosen range of R . Further, observe that the curvature of the upper bounds approximately follow the curvature of the simulated violation probability around its minimum value and only deviates at higher sampling rate. This is an interesting property as it suggests that a rate that minimizes the upper bound will be a ‘‘good’’ rate solution

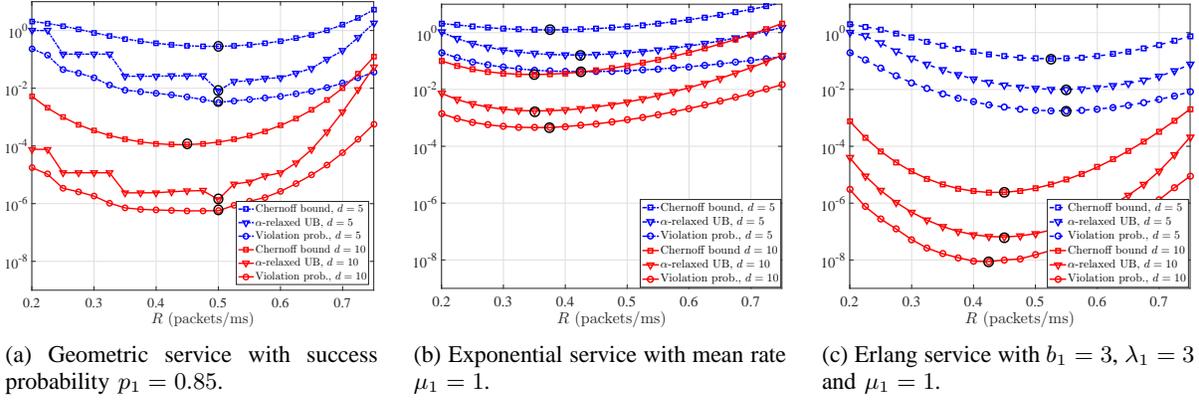


Fig. 4: Comparison of the upper bounds for varying arrival rate R in a *single hop* for different service time distributions.

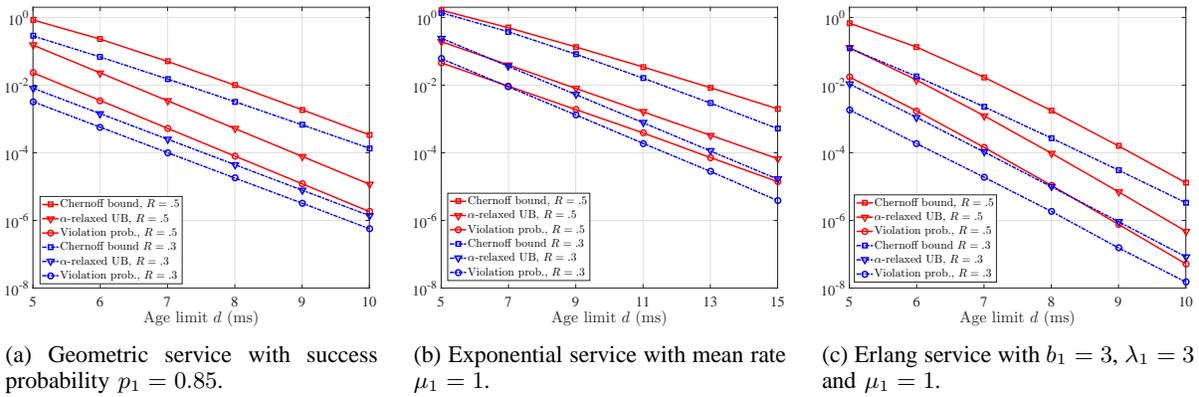


Fig. 5: Comparison of the upper bounds for varying age limit d in a *single hop* for different service time distributions.

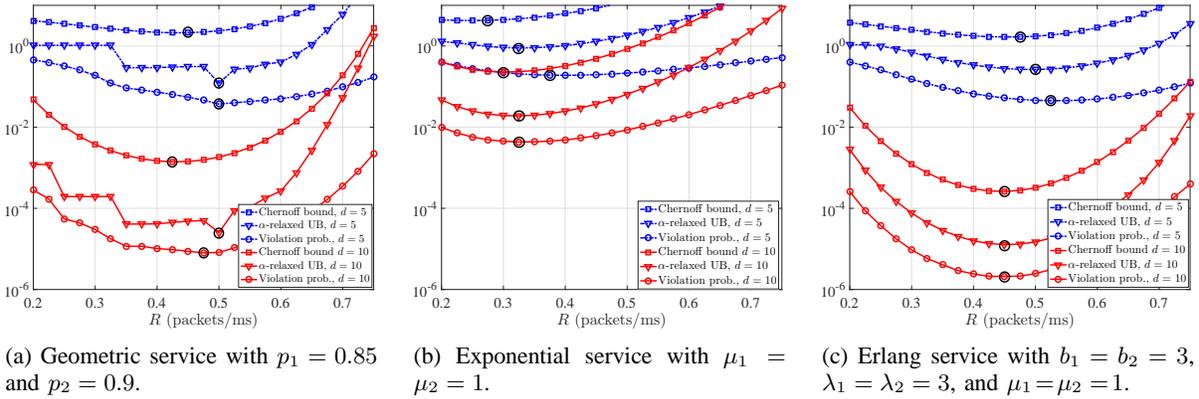


Fig. 6: Comparison of the upper bounds for varying arrival rate R in a *two hop* network for different service time distributions.

for minimizing the violation probability. We note that the α -relaxed upper bound curves are not continuous because the probability terms $\Phi(v_0, v_1, R)$ involves a floor function, namely, $\lfloor \beta \rfloor$. From Figure 5, we observe that the decay rates of the upper bounds match closely the decay rate of the violation probability. This further

strengthens our statement above that minimizing the upper bounds results in good heuristic rate solutions for the considered range of age limits.

2) *Two Hop*: In Figures 6 and 7, we present the upper bounds and the simulated AoI violation probability for varying arrival rate R and varying age limit d for

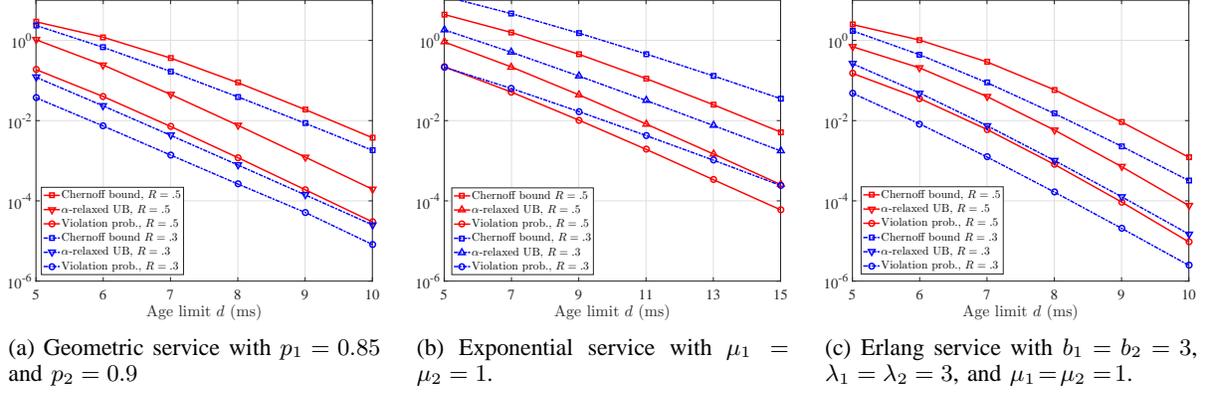


Fig. 7: Comparison of the upper bounds for varying age limit d in a *two hop* network for different service time distributions.

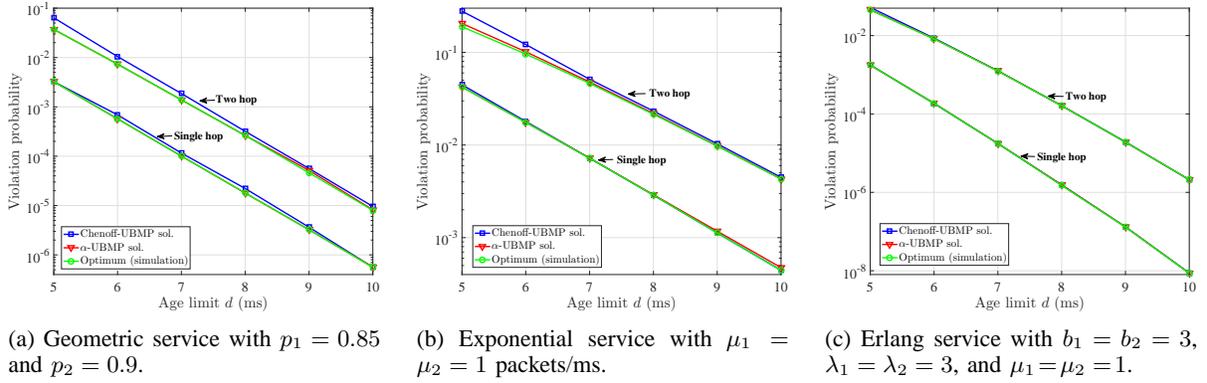


Fig. 8: Evaluation of the rate solutions obtained using upper bound minimization for different service time distributions.

different distributions for the two-hop scenario. We observe similar trends as in the case of single-hop scenario. Nevertheless, the bounds become relatively looser. This can be attributed to the fact that the union bound is applied twice for the two-hop scenario.

Note that for both single-hop and two-hop scenarios α -relaxed bound is much lower than the Chernoff bound. Nevertheless, Chernoff bound can be useful for the cases where the exact distribution of the summation of service times is intractable.

3) *Service Times with Higher Variance*: In this section, we study how the upper bounds perform for service time with higher variance. In Figure 9, we consider heterogeneous exponential service times, with $\mu_1 = 0.75$ and $\mu_2 = 1$. We have chosen $\mu_1 = 0.75$ so that the variance in this case is higher than that of the homogeneous case where both μ_1 and μ_2 are equal to 1. We note that the trend persists and that the mismatch in the minima among the three curves in this case becomes greater when lower age limit is considered, i.e., $d = 5$ ms. However, the mismatch is much smaller at higher d . We also note that compared to the homogeneous

server case in Figure 6(b), heterogeneity does not affect the conclusions regarding system behaviour with respect to AoI. Nevertheless, the performance becomes more dependent on the bottleneck link in this case.

In Figure 10, we consider hyper-exponential service time distribution with probability density function given by $p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x}$. We choose $p = 0.91$, $\lambda_1 = 0.95$, and $\lambda_2 = 2$ such that the mean value is equal to 1 ms. We note that this distribution has higher variance compared to exponential-service time distribution with mean 1. For computing the alpha-relaxed upper bound, we set $K = 6$ and numerically evaluated the convolution of hyper-exponential probability distribution functions to obtain values for $\phi(v_0, v_1, R)$.

From both Figures 9 and 10, we observe that for the two-hop scenario, for $d = 5$, the mismatch between the heuristic rate solution provided by α -UBMP and the optimal rate solution is relatively bigger. Nevertheless, under these settings, it should be noted that the value of the minimum AoI violation probability is not significantly lower than that achieved by the heuristic rate solution. Again, the main trends noticed with the other

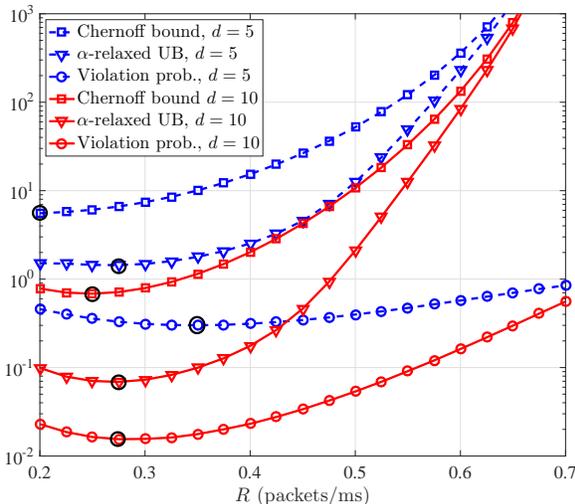


Fig. 9: Comparison of the upper bounds for varying arrival rate R in a *two hop* network for heterogeneous exponential service time distributions, with $\mu_1 = 0.75$ packet/ms and $\mu_2 = 1$ packets/ms.

three service distributions, see Figure 6, are prevailing here as well.

B. Quality of the Heuristic Solution

In Figure 8, we compare the violation probabilities for rate solutions obtained by solving the UBMPs and the estimated minimum/optimum violation probability obtained by exhaustive search using simulation, for both single-hop and two-hop scenarios. Note that the difference between the violation probabilities achieved by the heuristic rate solutions and the optimum violation probability is negligible. This suggests that the solutions of the UBMPs are near optimal for \mathcal{P} for the considered service-time distributions. This can be attributed to the fact that the upper bounds have decay rate that matches the decay rate of the violation probability as stated before. Although α -relaxed upper bound is much lower than Chernoff bound the solutions of α -UBMP provide only slightly lower violation probability than that of the Chernoff-UBMP solutions. Thus, Chernoff-UBMP is relatively tractable and the rate solutions provided can be used as first step toward computing close-to-optimal solutions by utilizing additional information about the service distributions.

Remark 3: We note that unlike the time-average age objective, which is minimized at 0.515 utilization factor (λ_1/μ_1) for the D/M/1 queue [4], the optimal rate solution and in turn the utilization factor that minimizes AoI violation probability depends on age limit d . For a comparison, in Figure 8(b) the single-hop scenario is equivalent to D/M/1 system

and in this case the optimal utilization factors are $\{0.425, 0.4, 0.4, 0.35, 0.35, 0.35, 0.35, 0.35\}$.

C. Queue Management Policies with Unit Buffer

Although this work is dedicated to study AoI distribution for WNCS under FCFS⁴ scheduling, recent research results have shown that considering a unit buffer with queue management policies provide lower AoI statistics in comparison to FCFS with infinite buffer [8], [10], [11], [23]. In this section, we investigate this effect by applying two widely referenced (in the context of AoI research) queue management policies, namely, FCFS-Unit Buffer and LGFS-Unit Buffer, to every queue in our network and then compare the achieved AoI violation probability to that we obtained earlier under FCFS. Both of these policies employs a one-packet buffer, however, they differ in that, whenever the buffer is occupied and a new packet arrives, the existing packet is kept in the first while it is replaced with the newly arriving packet in the second. In [8], it was shown that LGFS-Unit Buffer minimizes AoI processes, in stochastic ordering sense, among all non-preemptive service policies for any arrival process and service-time distributions. This implies, for the tandem two-queue system we consider, LGFS-Unit Buffer results in minimum AoI violation probability. Hence, it provides a good reference to measure the performance of other queue management policies against.

In Figures 11(a) and 11(b), the AoI violation probability is plotted against the arrival rate R for the FCFS as well as the two unit buffer queue management policies mentioned above, assuming exponential-service times with rate $\mu = 1$ packet/sec and for two age limits $d = \{5, 10\}$ ms. We observe that, the minimum AoI violation probability under FCFS-Unit Buffer and LGFS-Unit Buffer is comparable to that under FCFS in the single-hop scenario. However, the performance of FCFS-Unit Buffer deteriorates drastically in the two-hop case compared to the other two, see Figure 11(b). This can be attributed to the fact that, under FCFS-Unit Buffer, packets that are served at first link may still be dropped when arriving at the second link if its buffer is already occupied. This effect may be exasperated when more links (hops) are added to the tandem and even more when heterogeneous service processes is present at the different links along the cascade where latter links experience higher utilization. This example shows that in tandem-queuing systems, it is not always true that FCFS with finite buffer has lower AoI statistics than that of FCFS with infinite buffer. We contrast this with the performance trends of these policies for a system with parallel servers between the source and the destination [23], where it was demonstrated that,

⁴By FCFS we mean First-Come-First-Serve with infinite buffer.

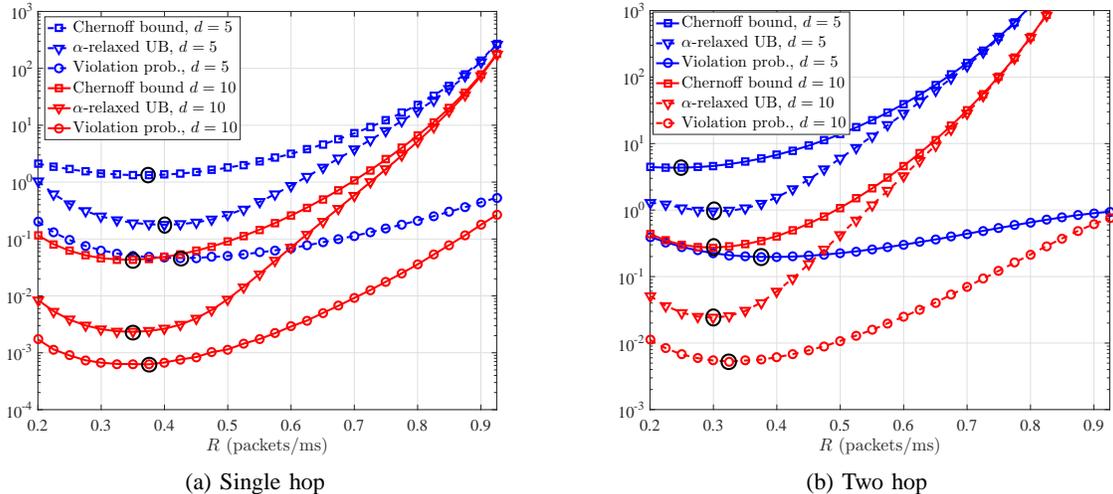


Fig. 10: Comparison of the upper bounds for varying arrival rate R for hyper-exponential service-time distribution.

under Poisson arrivals and exponential service-times, the minimum values for average AoI and average peak AoI achieved under FCFS with infinite buffer are much higher than that for the case with finite buffer.

Furthermore, in both scenarios, we observe that the AoI violation probability under FCFS with infinite queue is quite close to that of LGFS-Unit Buffer at low utilization (i.e., low R), and more importantly the minima for both cases are reasonably close and are achieved around the same arrival rate R . The reason for such behaviour can be attributed to the fact that at such low arrival rate the buffer would be empty most of the time mimicking the unit buffer behaviour. In the two-hop scenario, the FCFS would still have buffer space at the second hop to hold (and not drop as FCFS-Unit Buffer may do) a packet that is successfully forwarded by the first hop, hence, countering the performance deterioration experienced by FCFS-Unit Buffer due to packet drops that we highlighted in the two-hop scenario above.

The above observations are quite interesting. They suggest, at least for deterministic arrivals and tandem-queuing system, using FCFS with infinite queue may achieve a minimum AoI violation probability that is reasonably close to the optimum (achieved by LGFS-Unit Buffer among the set of non-preemptive policies [8]). This also opens an interesting research question for future work: how far the minimum AoI violation probability achieved under FCFS can be from the optimum?

IX. CONCLUSION AND FUTURE WORK

We provide a general characterization of AoI violation probability for a network with periodic input arrivals. Using this characterization, we formulate an optimization problem \mathcal{P} to find the optimal input rate which minimizes the AoI violation probability. Further, 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 computing an exact expression for the violation probability is hard, we propose an Upper Bound Minimization Problem (UBMP) and its more computationally tractable versions Chernoff-UBMP and α -UBMP, which result in heuristic rate solutions. We also present the Chernoff-UBMP for N-hop tandem queuing system. We solve Chernoff-UBMP and α -UBMP for single-hop and two-hop scenarios for three service-time distributions, namely, geometric, exponential and Erlang. Numerical results suggest that the rate solutions of α -UBMP are near optimal for \mathcal{P} , demonstrating the efficacy of our method. Furthermore, our simulation results suggest that, FCFS performs close to the optimum achieved by LGFS-Unit Buffer and drastically outperforms FCFS-Unit Buffer in a two-hop network with respect to AoI violation probability. This opens up an interesting research question: how far will be the minimum AoI violation probability achieved under FCFS from the optimum? Another interesting research direction for future work would be to extend 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. Finally, we would like to investigate different queuing disciplines relevant to AoI.

REFERENCES

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

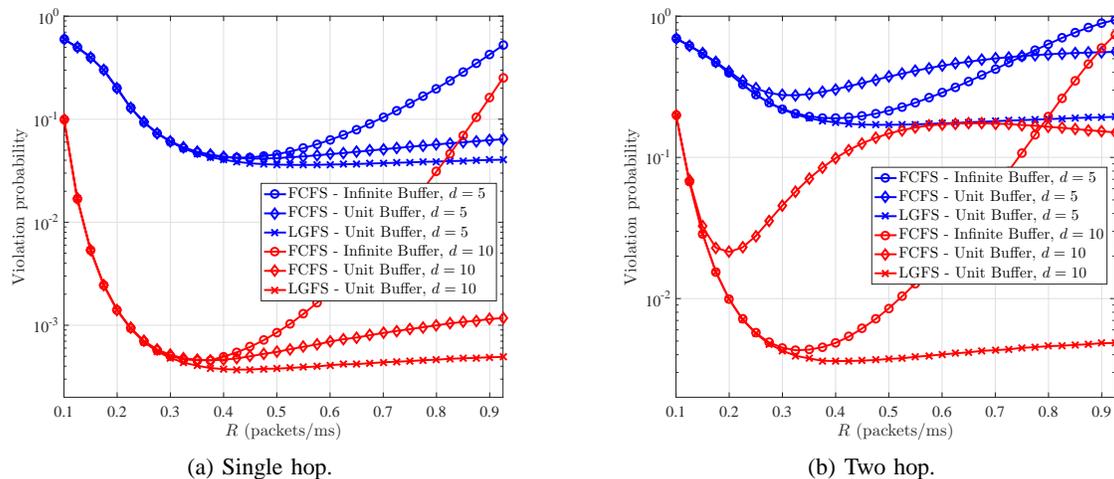


Fig. 11: Comparison of AoI violation probability achieved under different queue management policies under exponential-service times with rates equal to one packet/sec.

- [2] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for age of information for the D/G/1 queue," in *Proc. IEEE INFOCOM Workshop*, April 2018, pp. 130–135.
- [3] J. Hedberg, "Safety requirements specifications guideline," *SP Swedish National Testing and Research Institute*, 2005.
- [4] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012.
- [5] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE ISIT*, 2015.
- [6] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. ACM Mobihoc*, 2018, pp. 61–70.
- [7] S. Kaul, R. Yates, and M. Gruteser, "Status updates through queues," in *Proc. Conference on Information Sciences and Systems (CISS)*, 2012.
- [8] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, June 2017, pp. 576–580.
- [9] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," *CoRR*, vol. abs/1806.03487, 2018. [Online]. Available: <http://arxiv.org/abs/1806.03487>
- [10] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, April 2016.
- [11] 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 *Proc. IEEE INFOCOM*, April 2019, pp. 37–45.
- [12] K. Chen and L. Huang, "Age-of-information in the presence of error," *CoRR*, vol. abs/1605.00559, 2016.
- [13] E. Najm and R. Nasser, "Age of information: The gamma awakening," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2574–2578.
- [14] E. Najm, R. D. Yates, and E. Soljanin, "Status updates through M/G/1 queues with HARQ," in *Proc. IEEE ISIT*, June 2017, pp. 131–135.
- [15] A. Soysal and S. Ulukus, "Age of information in G/G/1 systems: Age expressions, bounds, special cases, and optimization," *CoRR*, vol. abs/1905.13743, 2019. [Online]. Available: <http://arxiv.org/abs/1905.13743>
- [16] 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.
- [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, E. Modiano, and R. Srikant, *Age of Information: A New Metric for Information Freshness*. Morgan & Claypool, 2019.
- [19] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE ISIT*, 2012.
- [20] R. Talak, S. Karaman, and E. Modiano, "Can determinacy minimize age of information?" *CoRR*, vol. abs/1810.04371, 2018. [Online]. Available: <http://arxiv.org/abs/1810.04371>
- [21] R. Talak and E. Modiano, "Age-delay tradeoffs in single server systems," in *Proc. IEEE ISIT*, July 2019, pp. 340–344.
- [22] R. Talak, S. Karaman, and E. Modiano, "When a heavy tailed service minimizes age of information," in *Proc. IEEE ISIT*, July 2019, pp. 345–349.
- [23] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5215–5232, 2019.
- [24] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE ISIT*, 2015.
- [25] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksul, 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.
- [26] J. P. Champati, R. R. Avula, T. J. Oechtering, and J. Gross, "On the minimum achievable age of information for general service-time distributions," July 2020, to be published in IEEE INFOCOM. [Online]. Available: <https://arxiv.org/abs/2001.06831>
- [27] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. Allerton Conference on Communication, Control, and Computing*, Oct 2017, pp. 486–493.
- [28] S. Farazi, A. G. Klein, and D. R. Brown, "Fundamental bounds on the age of information in general multi-hop interference networks," in *Proc. IEEE INFOCOM WKSHPS*, April 2019, pp. 96–101.
- [29] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in multihop multicast networks," *CoRR*, vol. abs/1812.10455, 2018. [Online]. Available: <http://arxiv.org/abs/1812.10455>
- [30] 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.
- [31] B. T. Bacinoglu, Y. Sun, E. Uysal, and V. Mutlu, "Optimal status updating with a finite-battery energy harvesting source," *Journal*

of *Communications and Networks*, vol. 21, no. 3, pp. 280–294, June 2019.

- [32] R. D. Yates, M. Tavan, Y. Hu, and D. Raychaudhuri, “Timely cloud gaming,” in *Proc. IEEE INFOCOM*, May 2017, pp. 1–9.
- [33] 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 Lemma 7

The violation probability is given by

$$\begin{aligned}
\mathbb{P}\{D(\hat{n}_R) > t\} &= \mathbb{P}\{D_2(\hat{n}_R) > t\} \\
&= \mathbb{P}\left\{\max_{0 \leq v_0 \leq \hat{n}_R} \left(A_2(\hat{n}_R - v_0) + \sum_{i=0}^{v_0} X_2^{\hat{n}_R - i}\right) > t\right\} \\
&= \mathbb{P}\left\{\bigcup_{v_0=0}^{\hat{n}_R} \left(A_2(\hat{n}_R - v_0) + \sum_{i=0}^{v_0} X_2^{\hat{n}_R - i} > t\right)\right\} \\
&\leq \sum_{v_0=0}^{\hat{n}_R} \mathbb{P}\left\{A_2(\hat{n}_R - v_0) + \sum_{i=0}^{v_0} X_2^{\hat{n}_R - i} > t\right\}. \quad (28)
\end{aligned}$$

Further, we have

$$\begin{aligned}
A_2(\hat{n}_R - v_0) &= D_1(\hat{n}_R - v_0) \\
&= \max_{0 \leq v_1 \leq \hat{n}_R - v_0} \left(A_1(\hat{n}_R - v_0 - v_1) + \sum_{i=0}^{v_1} X_1^{\hat{n}_R - v_0 - i}\right).
\end{aligned}$$

Substituting $A_2(\hat{n}_R - v_0)$ and $A_1(\hat{n}_R - v_0 - v_1) = \frac{\hat{n}_R - v_0 - v_1}{R}$ in (28) we obtain

$$\begin{aligned}
\mathbb{P}\{D(\hat{n}_R) > t\} &\leq \sum_{v_0=0}^{\hat{n}_R} \mathbb{P}\left[\max_{0 \leq v_1 \leq \hat{n}_R - v_0} \left(\frac{\hat{n}_R - v_0 - v_1}{R} + \sum_{i=0}^{v_1} X_1^{\hat{n}_R - v_0 - i}\right) + \sum_{i=0}^{v_0} X_2^{\hat{n}_R - i} > t\right] \\
&\leq \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \mathbb{P}\left[\frac{\hat{n}_R - v_0 - v_1}{R} + \sum_{i=0}^{v_1} X_1^{\hat{n}_R - v_0 - i} + \sum_{i=0}^{v_0} X_2^{\hat{n}_R - i} > t\right] \\
&= \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \mathbb{P}\left\{\sum_{i=0}^{v_1} X_1^i + \sum_{i=0}^{v_0} X_2^i > d + \frac{v_0 + v_1 - 1}{R}\right\}.
\end{aligned}$$

In the second step above, we have used the union bound. In the last step we have used $\hat{n}_R \leq R(t - d) + 1$. Also, since X_1 and X_2 are i.i.d., we re-indexed the superscripts of X_1 and X_2 in the summations. The result follows from the fact that as t goes to infinity \hat{n}_R goes to infinity.

B. Proof of Theorem 4

We first obtain Chernoff bound for $\Phi(v_0, v_1, R)$. We have

$$\begin{aligned}
\Phi(v_0, v_1, R) &= \mathbb{P}\left\{\sum_{i=0}^{v_1} X_1^i + \sum_{i=0}^{v_0} X_2^i > d + \frac{v_0 + v_1 - 1}{R}\right\} \\
&\leq \min_{s>0} e^{-s(d + \frac{v_0 + v_1 - 1}{R})} \mathbb{E}[e^{s(\sum_{i=0}^{v_1} X_1^i + \sum_{i=0}^{v_0} X_2^i)}] \\
&= \min_{s>0} e^{-s(d + \frac{v_0 + v_1 - 1}{R})} [M_1(s)]^{v_1 + 1} [M_2(s)]^{v_0 + 1} \\
&= \min_{s>0} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \beta_1^{v_1}(s) \beta_2^{v_0}(s). \quad (29)
\end{aligned}$$

Assuming the moment generating function of X exists, in the second step above we have used the Chernoff bound. In the third step above we have used the fact that X_n^k are i.i.d. for all k and n , and in the last step we have used (3). Using (29) in Lemma 7, we obtain

$$\begin{aligned}
\lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} &\leq \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \Phi(v_0, v_1, R) \\
&\leq \min_{s>0} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \phi(s, \beta_1(s), \beta_2(s)), \quad (30)
\end{aligned}$$

where

$$\phi(s, \beta_1(s), \beta_2(s)) = \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \beta_1^{v_1}(s) \beta_2^{v_0}(s). \quad (31)$$

Note that in the second step of (30) we have used the fact that for positive quantities sum over minimum is less than or equal to minimum over the sum. In the following lemma we provide a closed form expression for $\phi(s, \beta_1(s), \beta_2(s))$.

Lemma 11. For $s \in \mathcal{S}$,

$$\phi(s, \beta_1(s), \beta_2(s)) = \frac{1}{(1 - \beta_1(s))(1 - \beta_2(s))}.$$

Proof. Recall that $\beta_1(s) < 1$ and $\beta_2(s) < 1$, for all $s \in \mathcal{S}$. Using this, we obtain

$$\begin{aligned}
\phi(s, \beta_1^{v_1}, \beta_2^{v_0}) &= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \beta_1^{v_1}(s) \beta_2^{v_0}(s) \\
&= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \beta_2^{v_0}(s) \sum_{v_1=0}^{\hat{n}_R - v_0} \beta_1^{v_1}(s) \\
&= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \beta_2^{v_0} \cdot \frac{(1 - \beta_1^{\hat{n}_R - v_0 + 1}(s))}{1 - \beta_1(s)} \\
&= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \left[\frac{\beta_2^{v_0}}{1 - \beta_1(s)} - \frac{\beta_2^{v_0}(s) \beta_1^{\hat{n}_R - v_0 + 1}(s)}{1 - \beta_1(s)} \right] \\
&= \frac{1}{(1 - \beta_1(s))(1 - \beta_2(s))} - \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \frac{\beta_2^{v_0}(s) \beta_1^{\hat{n}_R - v_0 + 1}(s)}{1 - \beta_1(s)}
\end{aligned}$$

It is now sufficient to show that the summation term above is equal to zero. We first note that the summation is non-negative since $0 \leq \beta_1(s) < 1$ and $0 \leq \beta_2(s) < 1$. Let $\beta(s) = \min(\beta_1(s), \beta_2(s))$, then we have

$$\begin{aligned} \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \frac{\beta_2^{v_0}(s) \beta_1^{\hat{n}_R - v_0 + 1}(s)}{1 - \beta_1(s)} &\leq \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \frac{\beta^{\hat{n}_R + 1}(s)}{1 - \beta_1(s)} \\ &= \lim_{\hat{n}_R \rightarrow \infty} \frac{(\hat{n}_R + 1)}{\beta^{-(\hat{n}_R + 1)}(s)(1 - \beta_1(s))} \\ &= \lim_{\hat{n}_R \rightarrow \infty} \frac{1}{\beta^{-(\hat{n}_R + 1)}(s)(-\log \beta(s))(1 - \beta_1(s))} = 0. \end{aligned}$$

In the third step above we have used L'Hospital's Rule. Since the summation is non-negative and is less than or equal to zero, it should be equal to zero. \square

It is easy to see that if $s \notin \mathcal{S}$, then $\phi(s, \beta_1(s), \beta_2(s))$ will be equal to infinity. Therefore, using (30) and Lemma 11, we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} &\leq \min_{s > 0} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \phi(s, \beta_1(s), \beta_2(s)) \\ &= \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} \cdot \frac{M_1(s) M_2(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}. \end{aligned}$$

Hence the result is proven.

C. Proof of Theorem 5

From Lemma 7 we have

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{P}\{D(\hat{n}_R) > t\} &\leq \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \Phi(v_0, v_1, R) \\ &= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{K-1} \sum_{v_1=0}^{K-1} \Phi(v_0, v_1, R) + \Phi_1(K) + \Phi_2(K), \end{aligned} \quad (32)$$

where

$$\begin{aligned} \Phi_1(K) &= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{K-1} \sum_{v_1=K}^{\hat{n}_R - v_0} \Phi(v_0, v_1, R) \\ \Phi_2(K) &= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=K}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \Phi(v_0, v_1, R). \end{aligned}$$

In the following we use the Chernoff bound for $\Phi(v_0, v_1, R)$, given in (29), to derive bounds for $\Phi_1(K)$ and $\Phi_2(K)$. We have

$$\Phi_1(K) \leq \min_{s > 0} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \phi_1(s, \beta_1(s), \beta_2(s)), \quad (33)$$

where

$$\phi_1(s, \beta_1(s), \beta_2(s)) = \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{K-1} \sum_{v_1=0}^{\hat{n}_R - v_0} \beta_1^{v_1}(s) \beta_2^{v_0}(s)$$

$$\begin{aligned} &= \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{K-1} \beta_2^{v_0}(s) \frac{\beta_1^K(s)(1 - \beta_1^{\hat{n}_R - v_0 + 1}(s))}{1 - \beta_1(s)} \\ &= \frac{\sum_{v_0=0}^{K-1} \beta_2^{v_0}(s) \beta_1^K(s)}{1 - \beta_1(s)} - \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=0}^{K-1} \frac{\beta_2^{v_0}(s) \beta_1^{\hat{n}_R - v_0 + 1 + K}(s)}{1 - \beta_1(s)} \\ &= \frac{(1 - \beta_2^K(s)) \beta_1^K(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}, \text{ for } s \in \mathcal{S}. \end{aligned} \quad (34)$$

The second term in the third step above vanishes as $\beta_1(s) < 1$ for $s \in \mathcal{S}$. Using (34) in (33), we obtain

$$\Phi_1(K) \leq \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \frac{(1 - \beta_2^K(s)) \beta_1^K(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}. \quad (35)$$

Again, substituting (29) in $\Phi_2(K)$, we obtain

$$\Phi_2(K) \leq \min_{s > 0} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \phi_2(s, \beta_1(s), \beta_2(s)), \quad (36)$$

where

$$\phi_2(s, \beta_1(s), \beta_2(s)) = \lim_{\hat{n}_R \rightarrow \infty} \sum_{v_0=K}^{\hat{n}_R} \sum_{v_1=0}^{\hat{n}_R - v_0} \beta_1^{v_1}(s) \beta_2^{v_0}(s).$$

Using similar analysis as in Lemma 11, we obtain

$$\phi_2(s, \beta_1(s), \beta_2(s)) = \frac{\beta_2^K(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}, \text{ for } s \in \mathcal{S}. \quad (37)$$

Using (37) in (36), we obtain

$$\Phi_2(K) \leq \min_{s \in \mathcal{S}} e^{-s(d - \frac{1}{R})} M_1(s) M_2(s) \frac{\beta_2^K(s)}{(1 - \beta_1(s))(1 - \beta_2(s))}. \quad (38)$$

Finally, substituting (35) and (38) in (32) we obtain the result.



Jaya Prakash Champati is a post-doctoral researcher from the division of Information Science and Engineering, EECS, KTH Royal Institute of Technology, Sweden. He finished his PhD in Electrical and Computer Engineering, University of Toronto, Canada in 2017. He obtained his master of technology degree from Indian Institute of Technology (IIT) Bombay, India, and bachelor of technology degree from National Institute of Technology Warangal, India. His general research interest

is in the design and analysis of algorithms for scheduling problems that arise in networking and information systems. Currently, his focus is on freshness and delay analysis for time-critical control applications in Cyber-Physical Systems (CPS) and the Internet of Things (IoT). In the past, he worked on task scheduling and job assignment problems with motivations from computational offloading in edge computing systems. Prior to joining PhD he worked at Broadcom Communications, where he was involved in developing the LTE MAC layer. He was a recipient of the best paper award at IEEE National Conference on Communications, India, 2011.



Hussein Al-zubaidy (S07M11SM16) received the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2010. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, from 2011 to 2013. In the Fall of 2013, he joined the School of Electrical Engineering (EES) at the Royal Institute of Technology (KTH), Stockholm, Sweden, as a Post-Doctoral Fellow.

Since Fall 2015, he has been a Senior Researcher with EES at the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Al-Zubaidy is the recipient of many honors and awards, including the Ontario Graduate Scholarship (OGS), NSERC Visiting Fellowship, NSERC Summer Program in Taiwan, OGSST, and NSERC Post-Doctoral Fellowship.



James Gross received his Ph.D. degree from TU Berlin in 2006. From 2008-2012 he was Assistant Professor and head of the Mobile Network Performance Group at RWTH Aachen University, as well as a member of the DFG-funded UMIC Research Centre of RWTH. Since November 2012, he has been with the Electrical Engineering and Computer Science School, KTH Royal Institute of Technology, Stockholm, where he is professor for machine-to-machine communications.

He served as Director for the ACCESS Linnaeus Centre from 2016 to 2019, while he is currently a member of the board of KTHs Innovative Centre for Embedded Systems. His research interests are in the area of mobile systems and networks, with a focus on critical machine-to-machine communications, cellular networks, resource allocation, as well as performance evaluation methods. He has authored about 150 (peer-reviewed) papers in international journals and conferences. His work has been awarded multiple times, including the Best Paper Award at ACM MSWiM 2015, the Best Demo Paper Award at IEEE WoWMoM 2015, the Best Paper Award at IEEE WoWMoM 2009, and the Best Paper Award at European Wireless 2009. In 2007, he was the recipient of the ITG/KuVS dissertation award for his Ph.D. thesis. He is also co-founder of R3 Communications GmbH, a Berlin-based start-up in the area of ultrareliable low-latency wireless networking for industrial automation.