

Peak Age of Information Distribution Bounds for Multi-Connectivity Transmissions

Federico Chiariotti, Beatriz Soret, and Petar Popovski

Department of Electronic Systems, Aalborg University

Fredrik Bajers Vej 7C, 9220 Aalborg, Denmark, email: {fchi,bsa,petarp}@es.aau.dk

Abstract—Transmission of packets over multiple wireless interfaces is an effective method to improve reliability and reduce the delay of Internet of Things (IoT) transmissions. In this context, Age of Information (AoI) has become a useful metric for many applications, measuring the freshness of the information available on a process that is measured by a remote sensor. In this work, we study the Peak Age of Information (PAoI) of an $M/M/2$ fork-join system, in which a sensor sends packets simultaneously over 2 separate queuing systems. The first packet to reach the receiver is considered as delivered. We derive lower and upper bounds on the PAoI for systems with finite and infinite buffers based on a low-traffic approximation, and show that the bounds are very tight at the optimal working point, so that the best rate derived from the bounds is very close to the optimum.

Index Terms—Age of Information, multi-connectivity transmission, fork-join

I. INTRODUCTION

The new classes of communication in 5G and beyond networks, such as Ultra-Reliable Low-Latency Communication (URLLC) and massive Machine Type Communications (mMTC), have led to a redefinition of the objectives and metrics of communication systems. In particular, the stringent latency and reliability requirements of URLLC have led to the use of multi-connectivity, which can provide a higher reliability by exploiting path diversity. However, URLLC's requirements mean that packets cannot be too frequent, to avoid overusing the spectrum resources, so constant monitoring application cannot exploit its guaranteed latency. In this case, Age of Information (AoI) [1] is the metric of choice, as it can represent the freshness of the available information better than latency [2]. Specifically, we will concentrate on the Peak Age of Information (PAoI) [3], which measures age at the instant when each packet arrives.

The use of queuing theory for models with multiple parallel connections has been widely explored in the parallel computing literature: in particular, any system which has synchronized inputs to multiple queues is called a *fork-join* model in the relevant literature [4]. This kind of model is often used for parallel and distributed computing tasks, as well as for multipath communications [5]. Most of the works assume that, at any time instant, tasks can be canceled and abandon their respective queue. The trade-off between latency and computational cost is the standard parameter in the context of fork-join models, either for cloud computing [6] or multiple parallel transmissions [7]. Naturally, the more redundancy is added in the transmission or computation, the more the load on

the system will increase, so there is often an optimal amount of redundancy that depends on the capacity of the system and on the total load. In some systems, not adding any redundancy at all might be the optimal choice [8]. Some works have also considered worst-case performance, computing a bound of the tail of the latency distribution [9] with Markovian arrivals.

Our work also focuses on worst-case performance, but considering the PAoI metric: we compute lower and upper bounds of the PAoI distribution in $M/M/2$ fork-join systems, which are useful to understand the worst-case performance of duplicated transmissions. PAoI, and AoI in general, have become common metrics in networking due to their closer relation to monitoring processes than pure latency. They are relevant for all systems that require a fresh knowledge of a remote process in order to monitor or control it. Most works in the literature [10] concentrate on average age, both in simple queuing systems with a single node and unicast scheduled transmissions [11] and more complex and realistic wireless access scenarios such as multi-hop transmissions [12]. The number of works deriving higher moments or the full distribution of the age is much smaller [13], [14], although a reliable system design requires knowing the probability of occurrence of rare, but extremely damaging events.

More specifically, there are two papers that consider the AoI in fork-join systems, both of which consider the average age in systems with M parallel servers [5], [15]. They prove that a system designed to minimize the age will do it at the expenses of high waiting times and service times for the packets that do not contribute to the age metric (the *non-informative* packets) and the average and variance of the packet delay will therefore increase enormously. Unlike most of the fork-related works, we do not consider removal of tasks in the buffers, since age-sensitive applications will typically have no feedback.

II. SYSTEM MODEL AND PEAK AGE BOUNDS

A. Queueing system model

We consider a duplicated $M/M/2$ queuing system as shown in Fig. 1. Packets are generated according to a Poisson process with rate λ , and then duplicated and put in the queues for two different queuing systems with exponentially distributed service times with rates μ_1 and μ_2 , respectively. We also define $\alpha_j = \mu_j - \lambda_j$ $j \in \{1, 2\}$. The two queues have infinite capacity and follow a First Come First Serve (FCFS) queuing policy. The i -th packet is generated at time g_i , and received on each system at times $(r_{i,1}, r_{i,2})$. As the two copies of the packet

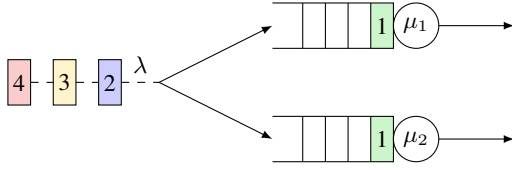


Fig. 1: Simple representation of the system model.

are redundant, we consider $r_i = \min(r_{i,1}, r_{i,2})$. We can then define the PAoI Δ_i as:

$$\Delta_i = r_i - g_{i-1}. \quad (1)$$

Its Probability Density Function (PDF) is denoted by $p_{\Delta_i}(\delta_i)$. We also define the latency of packet i on system j , $T_{i,j} = r_{i,j} - g_i$, with $T_i = r_i - g_i$, and the inter-arrival time $Y_i = g_i - g_{i-1}$. We then have that $\Delta_i = Y_i + T_i$.

Duplicated $M/M/2$ systems follow the Flatto-Hahn-Wright (FHW) model, which is known to be intractable; Flatto [16], [17] computed the generating function of the steady-state distribution of the number of packets in the two queues, but there is no closed-form solution for either the state or the waiting time distributions.

We consider a *lower bound* on the PAoI by using the steady-state probability distribution of the equivalent $M/M/2/L$ blocking fork-join queue. Naturally, this is a lower bound, because we do not consider the blocked packets, but the bound is tight if the blocking probability is low, i.e., if $P(q_{i,j} > L) \ll 1$, where $q_{i,j}$ is the number of queued packets that packet i finds when it arrives to system j . Under this condition the system almost never reaches a full queue and operates like the infinite buffer case. The *upper bound* is obtained by considering all cases in which the queue reaches L as a sample with infinite age. The same bounds are valid for both the $M/M/2/L$ and the $M/M/2/\infty$ fork-join system, and in both cases they are tighter if the traffic load is low.

B. Transition probabilities

The $M/M/2/L$ blocking queue can be represented by a finite semi-Markov chain with state $\mathbf{Q} = (Q_1, Q_2)$, where Q_j represents the number of packets currently in system j , and transition matrix \mathbf{P} . We have three possible transitions, with all other elements of the transition matrix being equal to 0. The first transition is the arrival of a new packet in the queues:

$$P_{\mathbf{q}, \mathbf{q}+(1,1)} = \begin{cases} 1, & \text{if } \mathbf{q} = (0, 0); \\ \frac{\lambda}{\lambda + \mu_2}, & \text{if } q_1 = 0, 0 < q_2 < L; \\ \frac{\lambda}{\lambda + \mu_1}, & \text{if } 0 < q_1 < L, q_2 = 0; \\ \frac{\lambda}{\lambda + \mu_1 + \mu_2}, & \text{if } 0 < q_1 < L, 0 < q_2 < L; \\ 0, & \text{if } q_1 = L \vee q_2 = L. \end{cases} \quad (2)$$

The second transition is a departure from the first system:

$$P_{\mathbf{q}, \mathbf{q}-(1,0)} = \begin{cases} 0, & \text{if } q_1 = 0; \\ \frac{\mu_1}{\lambda + \mu_1}, & \text{if } 0 < q_1 < L, q_2 = 0; \\ \frac{\mu_1}{\lambda + \mu_1 + \mu_2}, & \text{if } 0 < q_1 < L, 0 < q_2 < L; \\ \frac{\mu_1}{\mu_1 + \mu_2}, & \text{if } 0 < q_1 < L, q_2 = L; \\ \frac{\mu_1}{\mu_1 + \mu_2}, & \text{if } q_1 = L, 0 < q_2 < L; \\ 1, & \text{if } q_1 = L \wedge q_2 = 0. \end{cases} \quad (3)$$

The third transition is a departure from the second system, and has the same elements with inverted indices, i.e., considering $P_{\mathbf{q}, \mathbf{q}-(0,1)}$. This relies on the common assumption of non-simultaneous departures, and on the independence of the two systems' service times. We can then get the steady-state probability, which we denote as ϕ :

$$\begin{aligned} \phi(\mathbf{P} - \mathbf{I}) &= 0 \\ \sum_{q_{i,1}=0}^L \sum_{q_{i,2}=0}^L \phi(\mathbf{q}_i) &= 1 \end{aligned} \quad (4)$$

However, transition times between states are not uniform, so we need to compute the average time that the system spends in each state \mathbf{q}_i , which is given by:

$$\tau(\mathbf{q}) = \begin{cases} \frac{1}{\lambda}, & \text{if } \mathbf{q} = (0, 0); \\ \frac{1}{\lambda + \mu_2}, & \text{if } q_1 = 0, 0 < q_2 < L; \\ \frac{1}{\lambda + \mu_1}, & \text{if } 0 < q_1 < L, q_2 = 0; \\ \frac{1}{\lambda + \mu_1 + \mu_2}, & \text{if } 0 < q_1 < L, 0 < q_2 < L; \\ \frac{1}{\mu_2}, & \text{if } q_{i,1} = 0, q_2 = L; \\ \frac{1}{\mu_1}, & \text{if } q_{i,1} = L, q_2 = 0; \\ \frac{1}{\mu_1 + \mu_2}, & \text{if } 0 < q_1 \leq L, q_2 = L; \\ \frac{1}{\mu_1 + \mu_2}, & \text{if } q_1 = L, 0 < q_2 < L. \end{cases} \quad (5)$$

We can then weight the distribution ϕ to get the steady-state distribution π :

$$\pi(\mathbf{q}_i) = \frac{\phi(\mathbf{q}_i)\tau(\mathbf{q}_i)}{\sum_{\mathbf{q} \in \{0, \dots, L\}^2} \tau_{\mathbf{q}}}, \quad \forall \mathbf{q}_i \in \{0, \dots, L\}^2. \quad (6)$$

We can now examine the transition probabilities between two consecutive arrivals, which is necessary to compute the PAoI: we consider the system to be in steady state when packet $i-1$ arrives, and the PAoI is then the difference between the generation time g_{i-1} and the reception time r_i . In this case, we know by definition that there are no arrivals in the interval Y_i , so we get the following transition probabilities, considering only queue j :

$$P_{Q_{i,j}|Q_{i-1,j}, Y_i}(q_{i,j}|q_{i-1,j}y_i) = \begin{cases} \frac{(\mu_j y_i)^{q_{i-1,j}-q_{i,j}+1} e^{-\mu_j y_i}}{(q_{i-1,j}-q_{i,j}+1)!} & 0 < q_{i,j} \leq q_{i-1,j} + 1; \\ 1 - \sum_{n=0}^{q_{i-1,j}} \frac{(\mu_j y_i)^{q_{i-1,j}-n+1} e^{-\mu_j y_i}}{(q_{i-1,j}-n+1)!} & q_{i,j} = 0; \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where $j \in \{1, 2\}$. Since departures from the two queuing systems are independent, we get:

$$P_{\mathbf{Q}_i|\mathbf{Q}_{i-1}, Y_i}(\mathbf{q}_i|\mathbf{q}_{i-1}, y_i) = \prod_{j=1}^2 P_{Q_{i,j}|Q_{i-1,j}, Y_i}(q_{i,j}|q_{i-1,j} y_i) \quad (8)$$

We can now apply the law of total probability to remove the condition on Y_i , knowing that it is exponentially distributed, although we will not report the result here, as it will be implicit in the age bound and is quite cumbersome.

C. Auxiliary functions

The derivation of the bounds of the PAoI is supported by the definition of several auxiliary definitions that simplify the notation. First, the parameter $\psi_{i,j}$ is given by:

$$\psi_{i,j} = q_{i-1,j} - q_{i,j} + 1. \quad (9)$$

We also define $\psi_i = \psi_{i,1} + \psi_{i,2}$.

Finally, we define the auxiliary function $\theta_{m,n}^{(\lambda)}(a)$:

$$\begin{aligned} \theta_{m,n}^{(\lambda)}(a) &= \int_0^a e^{\lambda x} x^m (a-x)^n dx \\ &= \sum_{i=0}^n \binom{n}{i} \frac{(m+i)!(-1)^m a^{n-i}}{\lambda^{m+i+1}} \left(e^{\lambda a} \sum_{j=0}^{m+i} \frac{(-\lambda a)^j}{j!} - 1 \right). \end{aligned} \quad (10)$$

This auxiliary function is the solution to the following integral:

$$\begin{aligned} \int_0^a e^{\lambda x} x^m (a-x)^n dx &= \int_0^a \sum_{i=0}^n \binom{n}{i} (-1)^{i+1} e^{\lambda x} x^{m+i} a^{n-i} dx \\ &= \sum_{i=0}^n \binom{n}{i} \frac{(m+i)!(-1)^m a^{n-i}}{\lambda^{m+i+1}} \left(e^{\lambda a} \sum_{j=0}^{m+i} \frac{(-\lambda a)^j}{j!} - 1 \right). \end{aligned} \quad (11)$$

Another solution for negative values of λ is given in [18], using the Beta function $\beta(x, y)$ and the confluent hypergeometric function ${}_1F_1(a, b, z)$:

$$\theta_{m,n}^{(\lambda)}(a) = \beta(m+1, n+1) {}_1F_1(m+1, m+n+1, -\lambda a) a^{m+n+1}. \quad (12)$$

As this solution avoids summing large values with alternating signs, it is numerically more stable.

D. Lower bound

The lower bound is derived as follows. We highlight that we assume an error-free system, but the bound is also valid for a Packet Erasure Channel (PEC) in which some packets cannot be decoded, but still take the same transmission time. We divide the computation in four cases, in which the i -th packet finds both systems empty (case \mathcal{A}), only the first system empty (case \mathcal{B}), only the first system with a standing queue (case \mathcal{C}), or both systems full (case \mathcal{D}). First, we compute the probability of case \mathcal{A} , in which both queues are empty:

$$p(\mathcal{A}) = \pi(0, 0). \quad (13)$$

In this case, the lower bound on the conditioned PDF of the PAoI is given by:

$$\begin{aligned} p_{\Delta_i|\mathcal{A}}(\delta_i) &= \int_0^{\delta_i} p_{Y_i}(y) \sum_{q_{i-1,1}=0}^L u(\delta_i) \sum_{q_{i-1,2}=0}^L \frac{\pi(\mathbf{q}_{i-1})}{p(\mathcal{A})} \\ &\quad P_{\mathbf{Q}_i|\mathbf{Q}_{i-1}, Y_i}((0, 0)|\mathbf{q}_{i-1}, y) p_{T_i|\mathbf{Q}_i, Y_i}(t|(0, 0), y) dy \\ &= \sum_{q_{i-1,1}=0}^L u(\delta_i) \sum_{q_{i-1,2}=0}^L \frac{\pi(\mathbf{q}_{i-1})}{p(\mathcal{A})} \int_0^{\delta_i} \lambda e^{-\lambda y} \left(1 - \sum_{\ell=0}^{q_{i-1,1}} \frac{(\mu_1 y)^\ell e^{-\mu_1 y}}{\ell!} \right) \\ &\quad \times \left(1 - \sum_{\ell=0}^{q_{i-1,2}} \frac{(\mu_2 y)^\ell e^{-\mu_2 y}}{\ell!} \right) (\mu_1 + \mu_2) e^{-(\mu_1 + \mu_2)(\delta_i - y)} dy \\ &= u(\delta_i) \sum_{q_{i-1,1}=0}^L \frac{(\mu_1 + \mu_2) \lambda e^{-(\mu_1 + \mu_2)\delta_i}}{p(\mathcal{A})} \sum_{q_{i-1,2}=0}^L \pi(\mathbf{q}_{i-1}) \\ &\quad \left[\frac{(e^{(\mu_1 + \mu_2 - \lambda)\delta_i} - 1)}{\mu_1 + \mu_2 - \lambda} - \sum_{k=0}^{q_{i-1,1}} \frac{\mu_1^k}{k!} \theta_{k,0}^{(\alpha_2)}(\delta_i) \right. \\ &\quad \left. - \sum_{k=0}^{q_{i-1,2}} \frac{\mu_2^k}{k!} \left(\theta_{k,0}^{(\alpha_1)}(\delta_i) - \sum_{j=0}^{q_{i-1,1}} \frac{\mu_1^j \theta_{j+k,0}^{(-\lambda)}(\delta_i)}{j!} \right) \right]. \end{aligned} \quad (14)$$

Case \mathcal{B} happens with probability:

$$p(\mathcal{B}) = \sum_{q_{i,2}=1}^L \pi((0, q_{i,2})). \quad (15)$$

The conditioned PDF in this case is:

$$\begin{aligned} p_{\Delta_i|\mathcal{B}}(\delta_i) &= \sum_{q_{i-1,1}=0}^L u(\delta_i) \sum_{q_{i-1,2}=1}^L \frac{\pi(\mathbf{q}_{i-1})}{p(\mathcal{B})} \sum_{q_{i,2}=1}^{q_{i-1,2}+1} \int_0^{\delta_i} \lambda e^{-(\lambda + \mu_2)y} \\ &\quad \left(1 - \sum_{j=0}^{q_{i-1,1}} \frac{(\mu_1 y)^j e^{-\mu_1 y}}{j!} \right) \frac{(\mu_2 y)^{\psi_{i,2}}}{\psi_{i,2}!} e^{-(\mu_1 + \mu_2)(\delta_i - y)} \\ &\quad \times \left[\frac{\mu_2^{q_{i,2}+1} (\delta_i - y)^{q_{i,2}}}{q_{i,2}!} + \mu_1 \sum_{k=0}^{q_{i,2}} \frac{(\mu_2 (\delta_i - y))^k}{k!} \right] dy \\ &= \sum_{q_{i-1,1}=0}^L u(\delta_i) \sum_{q_{i-1,2}=1}^L \frac{\pi(\mathbf{q}_{i-1})}{p(\mathcal{B})} \sum_{q_{i,2}=1}^{q_{i-1,2}+1} \frac{\lambda e^{-(\mu_1 + \mu_2)\delta_i}}{\psi_{i,2}!} \left[\sum_{j=0}^{q_{i,2}} \frac{\mu_2^{\psi_{i,2}+j}}{j!} \right. \\ &\quad \times \mu_1 \left(\theta_{\psi_{i,2},j}^{(\alpha_1)}(\delta_i) - \sum_{k=0}^{q_{i-1,1}} \frac{\mu_1^k \theta_{\psi_{i,2}+k,j}^{(-\lambda)}(\delta_i)}{k!} \right) \\ &\quad \left. + \frac{\mu_2^{q_{i-1,2}+2}}{q_{i,2}!} \left(\theta_{\psi_{i,2},q_{i,2}}^{(\alpha_1)}(\delta_i) - \sum_{k=0}^{q_{i-1,1}} \frac{\mu_1^k \theta_{\psi_{i,2}+k,q_{i,2}}^{(-\lambda)}(\delta_i)}{k!} \right) \right]. \end{aligned} \quad (16)$$

For case \mathcal{C} , the probability and PDF are the same as for case \mathcal{B} , with inverted indices. The probability of being in this case is then:

$$p(\mathcal{C}) = \sum_{q_{i,1}=1}^L \pi((q_{i,1}, 0)). \quad (17)$$

The PDF of the PAoI in this case is:

$$p_{\Delta_i|C}(\delta_i) = \sum_{q_{i-1,1}=1}^L u(\delta_i) \sum_{q_{i-1,2}=0}^L \pi(\mathbf{q}_{i-1}) \sum_{q_{i,1}=1}^{q_{i-1,1}+1} \lambda \frac{e^{-(\mu_1+\mu_2)\delta_i}}{p(\mathcal{B})\psi_{i,1}!} \\ \times \left[\sum_{j=0}^{q_{i,1}} \frac{\mu_2 \mu_1^{\psi_{i,1}+j}}{j!} \left(\theta_{\psi_{i,1},j}^{(\alpha_2)}(\delta_i) - \sum_{k=0}^{q_{i-1,2}} \frac{\mu_2^k \theta_{\psi_{i,1}+k,j}^{(-\lambda)}(\delta_i)}{k!} \right) \right. \\ \left. + \frac{\mu_1^{q_{i-1,1}+2}}{q_{i,1}!} \left(\theta_{\psi_{i,1},q_{i,1}}^{(\alpha_2)}(\delta_i) - \sum_{k=0}^{q_{i-1,2}} \frac{\mu_2^k \theta_{\psi_{i,1}+k,q_{i,1}}^{(-\lambda)}(\delta_i)}{k!} \right) \right]. \quad (18)$$

Finally, we consider case \mathcal{D} , in which both queues have at least one packet. The probability of this happening is:

$$p(\mathcal{D}) = \sum_{q_{i,1}=1}^L \sum_{q_{i,2}=1}^L \pi(\mathbf{q}_i). \quad (19)$$

In this case, the lower bound on the conditioned PDF of the PAoI is given by:

$$p_{\Delta_i|\mathcal{D}}(\delta_i) = \sum_{q_{i-1,1}=0}^L u(\delta_i) \sum_{q_{i-1,2}=0}^L \frac{\lambda \pi(\mathbf{q}_{i-1})}{p(\mathcal{D})} \sum_{q_{i,1}=1}^{q_{i-1,1}+1} \sum_{q_{i,2}=1}^{q_{i-1,2}+1} \int_0^{\delta_i} y^{\psi_i} \\ \frac{\mu_1^{\psi_{i,1}} \mu_2^{\psi_{i,2}} e^{-(\mu_1+\mu_2)\delta_i - \lambda y}}{\psi_{i,1}! \psi_{i,2}!} \left[\sum_{k=0}^{q_{i,1}} \frac{\mu_1^k \mu_2^{q_{i,1}+1}}{k!} \right. \\ \left. \times (\delta_i - y)^{q_{i,2}+k} + \sum_{k=0}^{q_{i,2}} \frac{\mu_1^{q_{i,1}+1} \mu_2^k}{k!} (\delta_i - y)^{q_{i,1}+k} \right] dy \\ = \sum_{q_{i-1,1}=1}^L u(\delta_i) \sum_{q_{i-1,2}=1}^L e^{-(\mu_1+\mu_2)\delta_i} \sum_{q_{i,1}=1}^{q_{i-1,1}+1} \pi(\mathbf{q}_{i-1}) \sum_{q_{i,2}=1}^{q_{i-1,2}+1} \lambda \frac{\mu_1^{q_{i-1,1}+2}}{p(\mathcal{D})\psi_{i,1}! \psi_{i,2}!} \\ \times \mu_2^{q_{i-1,2}+2} \left[\sum_{k=0}^{q_{i,1}} \frac{\theta_{\psi_{i,1},q_{i,2}+k}^{(-\lambda)}(\delta_i)}{\mu_1^{q_{i,1}+1-k} q_{i,2}! k!} + \sum_{k=0}^{q_{i,2}} \frac{\theta_{\psi_{i,1},q_{i,1}+k}^{(-\lambda)}(\delta_i)}{\mu_2^{q_{i,2}+1-k} q_{i,1}! k!} \right]. \quad (20)$$

The unconditioned PDF of the lower bound is then simply:

$$p_{L,\Delta_i}(\delta_i) = \sum_{S \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}} p(S) p_{\Delta_i|S}(\delta_i). \quad (21)$$

E. Upper bound

The upper bound considers any case in which both queues are full as having infinite age, and it is written in terms of the lower bound as:

$$p_{U,\Delta_i}(\delta_i) = \begin{cases} p_{L,\Delta_i}(\delta_i) (1 - \pi(L, L)) & \delta_i < \infty; \\ \pi(L, L) & \delta_i = \infty. \end{cases} \quad (22)$$

III. SIMULATION RESULTS

In order to verify the tightness of the bounds, we ran a Monte Carlo simulation with 1 million packets. We discarded the first 1000 to ensure that the system was in steady state, and plotted the CDFs of the PAoI for a symmetrical system with $\mu_1 = \mu_2 = 1$ and an asymmetrical one with $\mu_1 = 1$ and $\mu_2 = 1.25$. We considered the arrival rate λ as the basic tuning parameter, and set $L = 10$ for the bounds. This

value is relatively low and allows us to keep a manageable computational complexity for the bound calculation, as the complexity of the bound computation is $O(L^5)$, even when reusing parts of the calculation, and can easily grow beyond manageable levels. We have also simulated the age for an $M/M/2/10$ blocking queue, but have not plotted it, as it is very close to the $M/M/2/\infty$ for all considered values.

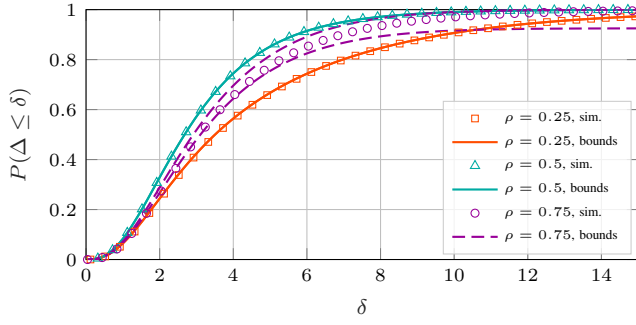
Fig. 2 shows the CDF of the peak age for three values of λ . As expected, transmitting either too often or too rarely has a negative impact on the PAoI. In the former case, the queues at the two system will tend to be longer, leading to a longer wait and, consequently, a higher age, while in the latter, the queues will be almost empty, but the inter-arrival times will become the main factor in the age. We can see this for the symmetrical system in Fig. 2a: the best performance is obtained with $\lambda = 0.5$, while the system has higher values of the PAoI for $\lambda = 0.25$ and $\lambda = 0.75$. As expected, our bounds are very tight for $\lambda = 0.25$ and $\lambda = 0.5$, with negligible difference from the empirical curve. As we increase the traffic load, the bounds start being looser. However, they are still relatively tight even for $\lambda = 0.75$. The lower bound is even closer in the asymmetrical case in Fig. 2b, almost matching the empirical distribution. In this case, the system performs equally well with the higher arrival rate, as one of the two systems can support it without increasing the queue significantly.

We can see this from Fig. 3, which shows the tail of the PAoI distribution, plotting its 95th and 99th percentiles as a function of λ . In the symmetrical system, the optimal λ is 0.58 for both percentiles, and the bounds remain extremely tight for λ up to 0.65 for the 99th percentile and 0.7 for the 95th. As we mentioned in the previous section, this is because the finite-size approximation becomes less adherent to reality as the probability of having queues larger than L increases. Naturally, setting a higher value of L improves the tightness of the bounds, but it also increases their computational complexity significantly. In the asymmetrical scenario, the optimal value of λ increases to 0.65, but so does the tightness of the bounds, which are tight past the optimal point. In these systems, even setting $L = 10$, which is a relatively low value and faster to compute, can enable system designers to optimize the arrival rate to minimize the PAoI.

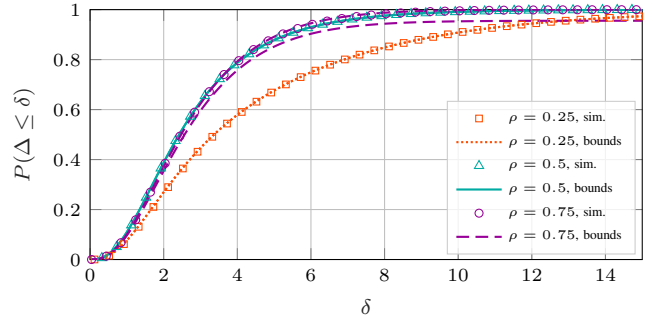
IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have derived bounds for the PAoI in $M/M/2$ fork-join systems, which can be used as a proxy for duplicated Internet of Things (IoT) transmissions. The bounds are computed by considering a finite-size queue with length L , and are a low-traffic approximation, extremely tight as long as the probability of going over the set queue size is small.

Future work involves the extension of the framework to multiple queues, which can model an arbitrary number of transmitting nodes, and other coding schemes such as K over N packet-level coding. We will also consider other kinds of fork-join queuing systems that can be relevant for other

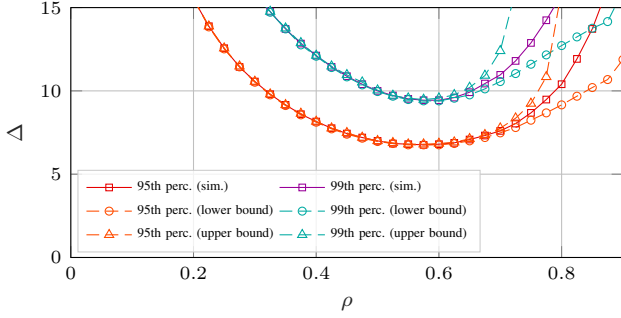


(a) $\mu_1 = \mu_2 = 1$, bounds with $L = 10$.

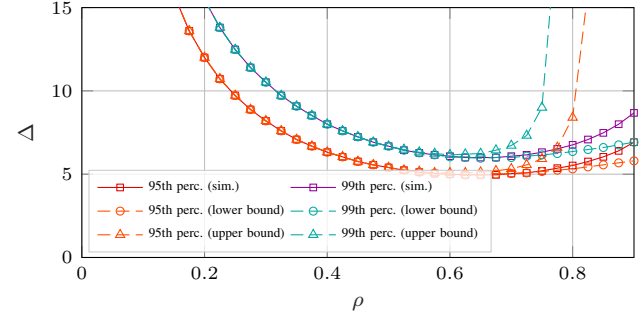


(b) $\mu_1 = 1, \mu_2 = 1.25$, bounds with $L = 10$.

Fig. 2: Cumulative Distribution Function (CDF) of the PAoI for different values of the system load ρ .



(a) $\mu_1 = \mu_2 = 1$, bounds with $L = 10$.



(b) $\mu_1 = 1, \mu_2 = 1.25$, bounds with $L = 10$.

Fig. 3: Tail of the PAoI distribution as a function of the system load ρ .

applications, such as the $D/M/2$ case, which can model the transmission of video frames over multiple paths.

ACKNOWLEDGMENT

This work is part of the IntelliIoT project, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 957218.

REFERENCES

- [1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *International Conference on Computer Communications (INFOCOM)*. IEEE, Mar. 2012, pp. 2731–2735.
- [2] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of Age of Information in the Internet of Things," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, Dec. 2019.
- [3] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *International Symposium on Information Theory (ISIT)*. IEEE, Jun. 2015, pp. 1681–1685.
- [4] C. Kim and A. K. Agrawala, "Analysis of the fork-join queue," *IEEE Transactions on Computers*, vol. 38, no. 2, pp. 250–255, Feb. 1989.
- [5] B. Buyukates and S. Ulukus, "Timely distributed computation with stragglers," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5273–5282, Jun. 2020.
- [6] G. Joshi, E. Soljanin, and G. Wornell, "Efficient redundancy techniques for latency reduction in cloud systems," *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS)*, vol. 2, no. 2, pp. 1–30, Feb. 2017.
- [7] N. B. Shah, K. Lee, and K. Ramchandran, "When do redundant requests reduce latency?" *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 715–722, Dec. 2015.
- [8] Y. Sun, C. E. Koksal, and N. Shroff, "On delay-optimal scheduling in queueing systems with replications," *ArXiv*, vol. abs/1603.07322, Mar. 2016.
- [9] A. Rizk, F. Poloczek, and F. Ciucu, "Stochastic bounds in fork-join queueing systems under full and partial mapping," *Queueing Systems*, vol. 83, no. 3, pp. 261–291, Aug. 2016.
- [10] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," *IEEE Transactions on Information Theory*, vol. 66, no. 9, pp. 5712–5728, May 2020.
- [11] A. Kosta, N. Pappas, V. Angelakis *et al.*, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, pp. 162–259, Nov. 2017.
- [12] N. Akar, O. Doğan, and E. U. Atay, "Finding the exact distribution of (Peak) Age of Information for queues of PH/PH/1/1 and M/PH/1/2 type," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5661–5672, Jun. 2020.
- [13] 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, Dec. 2019.
- [14] F. Chiarriotti, O. Vihrova, B. Soret, and P. Popovski, "Peak age of information distribution for edge computing with wireless links," *IEEE Transactions on Communications*, Jan. 2021.
- [15] R. Talak and E. H. Modiano, "Age-delay tradeoffs in queueing systems," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1743–1758, Mar. 2021.
- [16] L. Flatto and S. Hahn, "Two parallel queues created by arrivals with two demands I," *SIAM Journal on Applied Mathematics*, vol. 44, no. 5, pp. 1041–1053, Oct. 1984.
- [17] L. Flatto, "Two parallel queues created by arrivals with two demands II," *SIAM Journal on Applied Mathematics*, vol. 45, no. 5, pp. 861–878, Oct. 1985.
- [18] A. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and series, Volume 1: Elementary functions*. Gordon&Breach Scientific Publishing, New York, 1986.