Finding the Exact Distribution of (Peak) Age of Information for Queues of $PH/PH/1/1$ and $M/PH/1/2$ Type

Nail Akar, *Member, IEEE*, Ozancan Doğan, and Eray Unsal Atay

Abstract—**Bufferless and single-buffer queueing systems have recently been shown to be effective in coping with escalated Age of Information (AoI) figures arising in single-source status update systems with large buffers and FCFS scheduling. In this paper, for the single-source scenario, we propose a numerical algorithm for obtaining the exact distributions of both the AoI and the peak AoI (PAoI) in (i) the bufferless $PH/PH/1/1/P(p)$ queue with probabilistic preemption with preemption probability p, $0 \leq p \leq 1$, and (ii) the single buffer $M/PH/1/2/R(r)$ queue with probabilistic replacement of the packet in the queue by the new arrival with replacement probability r, $0 \leq r \leq 1$. The proposed exact models are based on the well-established theory of Markov Fluid Queues (MFQ) and the numerical algorithms are matrix-analytical and they rely on numerically stable and efficient vector-matrix operations. Moreover, the obtained exact distributions are in matrix exponential form, making it amenable to calculate the tail probabilities and the associated moments straightforwardly. Firstly, we validate the accuracy of the proposed method with simulations, and for sume sub-cases, with existing closed-form results. We then comparatively study the AoI performance of the queueing systems of interest under varying traffic parameters.**

Index Terms—**Age of Information, Peak Age of Information, PH-type distribution, Markov fluid queues.**

I. Introduction

CONSIDER the information update system in Fig. 1 consisting of one information source equipped with a sensor, a server local to the source in the form of a queue, and a remote monitor (or destination). The state of the information source is assumed to change in time which is detected by its sensor and the information source occasionally generates packets that contain sensed data along with a time stamp, to be immediately forwarded to the server. The server's role is to make decisions on when and which of these packets are to be sent towards the monitor which is responsible of collecting, monitoring, and further processing of these update messages. Packets are sent by the server to the monitor via a network which introduces random delays, i.e., service time of packets, and the monitor immediately sends back positive acknowledgments to the server. This single source model also studied in several other references including [1]–[4] is the focus of the current paper. The Age of Information (AoI) of a source is defined as the time elapsed since the generation of the last successfully received update packet at the monitor. The AoI concept was first introduced in [5] and later elaborated in [1], [6] as a metric to quantify the freshness of knowledge about the status of a remote information source in a status update system. In these studies, the update system described above is viewed as an abstraction of real-world scenarios and applications in which data freshness is crucial. There has recently been a surge of interest on AoI-related problems in various contexts including development of analytical models for AoI [4], [7], [8], AoI optimization methods [9]–[11], and AoI scheduling mechanisms [12], [13]. The reference [3] provides a relatively recent survey of the AoI concept and its applications. The focus of the current paper is on the development of the queueing models of the particular single-source, single-server scenario, and the scenarios involving multiple information sources sharing a single server [14], multiple servers [15], non-zero acknowledgment delays, and packet errors introduced by the network [8], are deliberately left outside the scope of this paper.

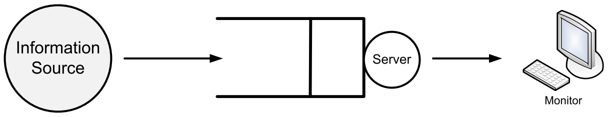

Fig. 1. An information source sending status update messages through a queue to a remote monitor.

The AoI processes of interest will first be introduced in the following very general setting not necessarily restricted to bufferless or single-buffer systems only. Using certain buffer management and scheduling mechanisms, the server is to send (in their entirety) a fraction of these packets, named successful packets, either immediately upon arrival or after a queue wait, towards the monitor. In this general description, some packets, named unsuccessful packets, are to be dropped at the server after receiving partial service or without receiving any service at all. Let $t_j (t'_j)$ denote the arrival instant of the j^{th},

Manuscript received October 23, 2019; revised April 12, 2020; accepted June 7, 2020. Date of publication June 17, 2020; date of current version September 16, 2020. The work of Ozancan Doğan was supported in part by the *5G and Beyond* Scholarship granted by the Information and Communication Technologies Authority (ICTA) of Turkey and Vodafone Turkey. The associate editor coordinating the review of this article and approving it for publication was A. Cohen. *(Corresponding author: Nail Akar.)*

The authors are with the Electrical and Electronics Engineering Department, Bilkent University, 06800 Ankara, Turkey (e-mail: akar@ee.bilkent.edu.tr).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.3002994

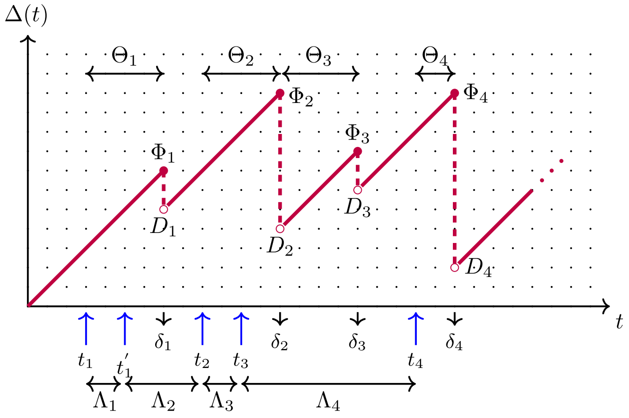

Fig. 2. Illustration of the sample path for the AoI process $\Delta(t)$ and the PAoI process Φ_j for a single-buffer queueing system with five information packets arriving at the server from the source at epochs 3, 5, 9, 11, and 20, respectively out of which the second packet is preempted by the third packet, all other four packets being successful. The successful packets 1 to 4 have a service time of 4, 4, 4, and 2, and a queue wait of 0, 0, 2, and 0, respectively. D_j is the system time, Θ_j is the service time, and δ_j is the reception time, of the $j^{\text{th}}, j = 1, \dots, 4$ successful packet, and Λ_i is the interarrival time between the packet arrivals (successful or not) i and $i+1$.

$j = 1, 2, \dots,$ successful (unsuccessful) information packet arriving at the queue. The interarrival time Λ between information packet (successful or unsuccessful) arrival epochs is assumed to be independent and identically distributed with cumulative distribution function (cdf) $F_\Lambda(\cdot)$, probability density function (pdf) $f_\Lambda(\cdot)$, mean $1/\lambda$, and squared coefficient of variation (scov) $c_\Lambda^2 = \sigma_\Lambda^2 \lambda^2$ where σ_Λ^2 denotes the variance of Λ. Let $\delta_j, j \geq 1$ denote the reception time at the monitor of the j^{th} successful packet. Also, let $D_j, j \geq 1$ denote the system time of successfully received packet j, which is the sum of two terms: i) the queue wait time W_j, and ii) the service time of the j^{th} received packet. Clearly, $D_j = \delta_j - t_j, j \geq 1$. In the current paper, all packet service times (denoted by Θ) are assumed to be independent and identically distributed with common cdf $F_\Theta(\cdot)$, pdf $f_\Theta(\cdot)$, mean $1/\mu$ and scov $c_\Theta^2 = \sigma_\Theta^2 \mu^2$ where σ_Θ^2 denotes the variance of Θ. The system load ρ is defined as $\rho = \lambda/\mu$.

Let $\Delta(t), t \geq 0$, denote the continuous-time random process with left-continuous sample paths representing the AoI for this source at time t with given $\Delta(0)$. After $t = 0$, $\Delta(t)$ increases linearly in time with a unit slope until the first packet reception at $t = \delta_1$. The right limit $\Delta(\delta_1^+) = \lim_{t \downarrow \delta_1} \Delta(t)$ is set to D_1 after which the process $\Delta(t)$ subsequently increases linearly in time with unit slope until the next reception and the pattern repeats forever. Let $\Phi_j = \Delta(\delta_j), j \geq 1$, denote the discrete-time continuous-valued random process (called the PAoI process) associated with the AoI just at the epoch of packet receptions. Fig. 2 shows a sample path of the random process $\Delta(t)$ for this general setting with $\Delta(0) = 0$ with five information packets arriving at a single buffer queue out of which the second packet is preempted by the third one. What is important to observe from Fig. 2 is that each cycle of the AoI process consists of a linear curve that starts at value D_j for some index j and increases at a unit rate until the peak value Φ_{j+1} of that cycle and the sample path of the AoI process consists of an ordered concatenation of infinitely many cycles. The difficulty of building analytical models for these two AoI processes stems from the fact that the random variables Φ_j and D_j are not independent and conventional queueing models do not cope well with jumps of this nature. For example, Markov Fluid Queues (MFQ) [16] that are instrumental to the current paper, can not produce sample paths given as in Fig. 2 due to the above-mentioned difficulty. However, as to be shown in the sequel, MFQs can be devised to produce sample paths that contain sample cycles whose parts coincide with the sample cycles of the AoI process while losing the ordering relationship of the AoI cycles within a sample path. Moreover, these cycles contain sample values that coincide with the sample values of the PAoI process, again losing the ordering relationship among PAoI samples. In the current paper, we show that such ordering is not needed for the marginal distributions of the AoI and PAoI processes and it is this observation that leads us to use the well-established tool of MFQs for AoI and PAoI analysis.

Depending on the distribution of the interarrival time Λ, the service time Θ, the queue capacity which represents the maximum number of packets that are allowed to be in the system (waiting room plus service), the scheduling discipline (First Come First Serve (FCFS), Preemptive Last Come First Serve (P-LCFS), Non-preemptive LCFS (NP-LCFS), etc.), and the buffer management scheme which refers to the specific decision of which packets are to be dropped at the server, there are several variations of the queueing system described above that have been studied in the recent literature. In most previous studies on AoI analysis, the focus has been on obtaining the mean AoI and mean PAoI values but there is also a need to obtain the distributions of the AoI and PAoI processes since their tail behaviors may also be important for the underlying status update system. In [1], the mean AoI is obtained for the $M/M/1$, $M/D/1$, and $D/M/1$ queues with infinite buffer capacity and FCFS scheduling. The authors of [17] derive expressions for the Laplace-Stieltjes transform of the stationary distributions of the AoI and the PAoI in $M/GI/1$ and $GI/M/1$ queues. However, these infinite buffer queueing systems that are quite popular in other contexts tend to perform quite poorly in terms of AoI in moderate to high load regimes. Recognizing this fact, the reference [7] studies the AoI and PAoI distributions in the $M/M/1/1$ and $M/M/1/2$ queues in which packets are dropped when the buffer is full at arrival epochs. The authors of [7] also introduce an alternative model, the so-called $M/M/1/2^*$ queue, for which the packet waiting in the queue is to be replaced by a new packet arrival, i.e., more formally named as $M/M/1/2/\mathit{NP\text{-}LCFS}$ queue. This particular system is further studied in [18] as well as the $M/M/1/1/\mathit{P\text{-}LCFS}$ preemptive queue where a new arrival preempts the packet in service and the service time distribution is assumed to follow a gamma distribution and mean AoI and mean PAoI results are given. The reference [4] presents exact expressions for the stationary distributions of AoI and PAoI for a very wide class of single-source information update systems including bufferless and single buffer variations. There are also several recent studies that attempt to obtain easily computable bounds for the crucial AoI-related metrics of interest. As an example, [19] derives exact expressions and upper bounds for the mean AoI for $G/G/1/1$ and $G/G/1/1/P\text{-}LCFS$ queues and they report that the upper bounds are in general close to exact average age expressions. Similarly, the authors of [20] present a simple methodology for obtaining upper bounds for the AoI violation probability for both $GI/GI/1/1$ and $GI/GI/1/2^*$ systems, in addition to some exact closed-form expressions for some sub-cases.

A random variable (rv) corresponding to the time until absorption of a continuous-time Markov chain with one absorbing state is said to possess a phase-type (PH-type or PH) distribution; see [21] for properties of PH-type distributions and their applications to performance modeling. Exponential, hyper-exponential, Erlang, and Coxian distributions, and their mixtures, are well-known examples of PH-type distributions. One important property is that the set of PH distributions is dense in the field of all positive-valued distributions and PH distributions can therefore be used to approximate any positive-valued distribution [22]. An Expectation Maximization (EM) algorithm for maximum likelihood estimation from sample data and density, for the purpose of approximation using PH distributions, is presented in [23] and very good fits to densities including Weibull, lognormal, etc. are obtained. The most well-known attribute of PH distributions is that problems arising in queueing systems that have an explicit solution assuming exponential distributions turn out to stay algorithmically tractable in case the exponential distribution is to be replaced with a PH distribution; see [23] and the references therein. In this way, using PH distributions allows one to algorithmically study the impact of higher order moments (such as the scov) of interarrival and service times on the performance of the queueing systems of interest.

In this paper, we propose to use PH distributions to generalize the existing results on AoI for modeling interarrival and service times. In particular, we propose a numerical algorithm for obtaining the exact distributions of both the AoI and PAoI processes as well as their moments for the following two systems:

(i) The bufferless $PH/PH/1/1/P(p)$ queue with PH-type information packet arrivals and PH-type service times, and a packet in service is preempted by a new arrival with preemption probability p, $0 \leq p \leq 1$. When $p = 0$, this model reduces to the ordinary $PH/PH/1/1$ queue, whereas for $p = 1$, the model is referred to as the LCFS preemptive $PH/PH/1/1/P\text{-}LCFS$ queue, or shortly the $PH/PH/1/1^*$ queue.

(ii) The single buffer $M/PH/1/2/R(r)$ queue with Poisson arrivals and PH-type service times and with a waiting room of one packet only, and the packet in the waiting room is replaced with packet replacement probability r, $0 \leq r \leq 1$. When $r = 0$, this model reduces to the $M/PH/1/2$ FCFS queue, whereas for $r = 1$, the model is referred to as the LCFS non-preemptive $M/PH/1/2/NP\text{-}LCFS$ queue, or shortly the $M/PH/1/2^*$ queue which is named in [7].

For both queueing systems, we are interested in finding the following steady-state cdfs for the random processes $\Delta(t)$, $t \geq 0$, and $\Phi_j, j \geq 1$:

$$F_\Delta(x) = \lim_{t\to\infty} \Pr\{\Delta(t) \leq x\}, F_\Phi(x) = \lim_{j\to\infty} \Pr\{\Phi_j \leq x\}, \tag{1}$$

for $x \geq 0$. Note that $F_\Delta(0)$ and $F_\Phi(0)$ must be zero since there can not be a probability mass at the origin for these two processes. Also, let $f_\Delta(x)$ and $f_\Phi(x)$ for $x \geq 0$ denote the corresponding steady-state pdfs. We have the following main contributions:

- For the analytic modeling of AoI and PAoI, we propose to use a generalization of the well-established theory of MFQs whose mathematical foundation goes back to the works [16], [24], [25]. Existing MFQ solvers that we use are matrix analytical and they rely on numerically stable and efficient vector-matrix operations. Moreover, the form of the obtained AoI and PAoI distributions is matrix exponential, making it amenable to calculate the tail probabilities, moments, etc. quite straightforwardly. This is in contrast with the need to invert Laplace transforms for finding age violation probabilities and the need for differentiating transforms for obtaining higher order moments as in [4].
- To the best of our knowledge, probabilistic preemption and replacement have not been analytically studied in the context of AoI distributions before. In this paper, we analytically model these probabilistic schemes and also demonstrate the merits of using probabilistic preemption in bufferless scenarios. Moreover, we have been able to provide a unifying algorithm for each of the bufferless and single-buffer queues through the probabilistic preemption and probabilistic replacement parameters.

The organization of the paper is as follows. In Section 2, preliminaries on PH distributions and MFQs are presented. Section 3 presents the proposed method for the two queueing systems of interest. In Section 4, we provide numerical examples to validate the proposed approach as well as the comparative assessment of the systems of interest under varying traffic parameters. Finally, we conclude in Section 5.

II. Preliminaries

A. Notation

Uppercase letters are used to denote real-valued matrices. Lowercase bold letters or symbols are used to denote real-valued vectors or scalars. The $(i, j)^{\text{th}}$ entry of A is denoted by $A_{i,j}$ and the j^{th} entry of a row or column vector $\boldsymbol{\alpha}$ is α_j. The notations $\mathbf{0}_{k\times\ell}$, $\boldsymbol{I}_m$, and $\mathbf{1}_n$ are used to denote the matrix of zeros of size $k \times l$, identity matrix of size m, and a column matrix of ones of size n, respectively. The subscripts are dropped when the sizes are clear from the context. Let A be an $n \times m$ matrix and B a $p \times q$ matrix. The Kronecker product of the matrices A and B is denoted by $A \otimes B$ which is of size $np \times mq$. The notation $\mathbf{diag}\{A, B\}$ denotes the block diagonal concatenation of the matrices A and B and is diagonal if the individual matrices A and B are diagonal. A square matrix is said to be stable (anti-stable) if each of its eigenvalues has negative (non-negative) real parts. The notation $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ is used for the concatenation of the two row vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. $u(x)$ denotes the Heaviside step function, also known as the unit step function. $\delta(x)$ denotes the Dirac delta function, also known as the unit impulse function.

B. Phase-Type Distributions

To describe a PH-type distribution, a Markov process is defined on the state-space $\mathcal{S} = \{1, 2, \ldots, m, m+1\}$ with one absorbing state $m+1$, initial probability vector $[\boldsymbol{\sigma}, \sigma_0]$, and an infinitesimal generator of the form $\left[\begin{array}{c:c} S & \boldsymbol{\nu} \\ \hdashline \mathbf{0} & 0 \end{array}\right]$, where $\boldsymbol{\sigma}$ is a row vector of size m, $\sigma_0 = 1 - \boldsymbol{\sigma}\mathbf{1}$ is a scalar, the subgenerator S is $m \times m$, and $\boldsymbol{\nu}$ is a column vector of size m such that $\boldsymbol{\nu} = -S\mathbf{1}$. The distribution of the time till absorption into the absorbing state $m + 1$, denoted by the random variable X, is called PH-type characterized with the pair $(\boldsymbol{\sigma}, S)$, i.e., $X \sim PH(\boldsymbol{\sigma}, S)$. Very commonly, the probability mass at zero vanishes for PH-type distributions used for modeling interarrival times and service times, i.e., $\sigma_0 = 0$ [21]. Note that $[\boldsymbol{\sigma}, \sigma_0]$ is a probability vector and the diagonal elements of S are strictly negative, its off-diagonal elements are non-negative and $S\mathbf{1} \leq \mathbf{0}$ elementwise. The cdf and the pdf of $X \sim PH(\boldsymbol{\sigma}, S)$, denoted by $F_X(x)$ and $f_X(x)$, respectively, are given as:

$$\begin{aligned} F_X(x) &= (1 - \boldsymbol{\sigma} e^{Sx}\mathbf{1})u(x), \\ f_X(x) &= -\boldsymbol{\sigma} e^{Sx} S\mathbf{1}u(x) + \sigma_0\delta(x). \end{aligned} \tag{2}$$

A generalization of the PH-type distribution is the so-called Matrix Exponential (ME) distribution [26], [27]. We say $X \sim ME(\boldsymbol{\sigma}, S)$ with order m when the pdf of the random variable X is in the same form (2) as in PH-type distributions, however, for the ME distribution, the parameters $\boldsymbol{\sigma}$ and S do not necessarily have the same stochastic interpretation. The moments of $X \sim PH(\boldsymbol{\sigma}, S)$ or $X \sim ME(\boldsymbol{\sigma}, S)$ are easy to obtain:

$$E[X^i] = i!\boldsymbol{\sigma}(-S)^{-i}\mathbf{1}, \quad i = 1, 2, \ldots. \tag{3}$$

Despite the lack of stochastic interpretation of the row vector $\boldsymbol{\sigma}$ and the matrix S, ME distributions can be substituted in place of PH-type distributions in computational techniques and algorithms [28]. We will also carry out a similar approach in this paper and use the more general ME distributions in place of PH-type distributions when necessary.

Lemma 1: Let the rv X have a pdf $f_X(x)$ in the following matrix exponential form:

$$\begin{aligned} f_X(x) &= \boldsymbol{g} e^{Ax}\boldsymbol{h}u(x) + \sigma_0\delta(x), \\ E[X^i] &= (-1)^{i+1} i! \boldsymbol{g} A^{-(i+1)}\boldsymbol{h}, \end{aligned} \tag{4}$$

for $i = 1, 2, \ldots,$ with A being of size n and $\int_{-\infty}^{\infty} f_X(x)\, dx = -\boldsymbol{g}A^{-1}\boldsymbol{h} + \sigma_0 = 1$. Then, X is ME-distributed, i.e., there exists a vector $\boldsymbol{\sigma}$ and a matrix S such that $X \sim ME(\boldsymbol{\sigma}, S)$ with order n.

Proof: When $h \neq 0$, there exists a nonsingular matrix M such that $M\mathbf{1} = \boldsymbol{v}$ where $\boldsymbol{v} = A^{-1}\boldsymbol{h}$. To show this, note that the eigenvalues of A need to be in the open left half plane for X to have a legitimate pdf. In particular, A is non-singular and therefore $\boldsymbol{v}$ has at least one non-zero element. Let k be the smallest index k such that $v_k \neq 0$. We construct the matrix M as follows. We set $M_{k,k} = v_k$ and for all $i \neq k$, if $v_i \neq 0$, we set $M_{i,i} = v_i$, and if $v_i = 0$, we set $M_{i,i} = 1, M_{i,k} = -1$. All the remaining entries of M are zero. Clearly, the matrix M constructed as above is non-singular and satisfies $M\mathbf{1} = \boldsymbol{v}$. Subsequently, one can easily show that $f_X(x)$ can be written in the form (2) with the choices of $\boldsymbol{\sigma} = \boldsymbol{g}M,\ S = M^{-1}AM$ and therefore $X \sim ME(\boldsymbol{\sigma}, S)$. ■

C. Markov Fluid Queues

A Markov Fluid Queue (MFQ) is described by a joint Markovian process $\boldsymbol{X}(t) = (X_f(t), X_m(t)), t \geq 0$, where $0 \leq X_f(t) < \infty, X_m(t) \in \mathcal{S} = \{1, 2, \ldots, n\}$, $X_f(t)$ represents the continuous-valued fluid level in the buffer and the modulating phase process $X_m(t)$ is a Continuous Time Markov Chain (CTMC) with state space $\mathcal{S}$ and generator Q. In MFQs, the net rate of fluid change (or drift) is r_i when the phase of the modulating process $X_m(t)$ is i. The drift matrix R is the diagonal matrix of drifts: $R = \mathbf{diag}\{r_1, r_2, \ldots, r_n\}$. When $X_f(t) = 0$ and $X_m(t) = i$ with $r_i < 0$, $X_f(t)$ sticks to the boundary at zero. The process $\boldsymbol{X}(t)$ is said to be characterized with the matrix pair (Q, R), i.e., $\boldsymbol{X}(t) \sim MFQ(Q, R)$. Stationary solution of infinite MFQs are studied in [16], [24] by using the eigendecomposition of a certain matrix. The reference [29] obtains the stationary solution of infinite and finite MFQs without having to find the eigenvectors, a problem which is known to be ill-conditioned [30].

The MFQ of interest to this paper is slightly different in the sense that $X_m(t)$ behaves according to generator Q as before when $X_f(t) > 0$ but it behaves according to another generator $\tilde{Q}$ when $X_f(t) = 0$. This generalized MFQ, called a GMFQ throughout this paper, is characterized with the ordered triple $(Q, \tilde{Q}, R)$, i.e., $\boldsymbol{X}(t) \sim GMFQ(Q, \tilde{Q}, R)$. The size of these matrices, n, gives the order of the GMFQ. GMFQs turn out to be a special case of the more general multi-regime MFQs studied in [31]. Therefore, their steady-state solutions are known and the underlying numerical methods are shown to be numerically stable and efficient. We assume $r_i \neq 0$, $1 \leq i \leq n$ which suffices for the AoI models developed in this paper. We further assume without loss of generality that there are b (a) states with positive (negative) drifts with $a + b = n$ and $r_i > 0,\ i \leq b$ and $r_i < 0,\ i > b$, since otherwise states can always be reordered for this purpose. For GMFQs, it is of interest the following steady-state joint pdf vector

$$\boldsymbol{f}(x) = [f_1(x), f_2(x), \ldots, f_n(x)], \tag{5}$$

where $f_i(x) = \lim_{t\to\infty} \frac{d}{dx} \Pr\{X_f(t) \leq x, X_m(t) = i\}$ for $x > 0$ and the steady-state probability mass accumulation (pma) vector at zero:

$$\boldsymbol{c} = [c_1, c_2, \ldots, c_n], \tag{6}$$

where $c_i = \lim_{t\to\infty} \Pr\{X_f(t) = 0, X_m(t) = i\}$. Alg. 1 describes the pseudo-code of obtaining the quantities of interest in (5) and (6) on the basis of the numerical algorithm proposed in [31] for more general multi-regime MFQs in cases when the steady-solution exists. The computational complexity of

Algorithm 1 Steady-State Solution of $GFMQ(Q, \tilde{Q}, R)$

```
function STEADY-STATE($Q, \tilde{Q}, R, n, a, b$)  ▷ $n$ is the size of these matrices
    Step 1: Find an orthogonal matrix $P$ that puts the matrix $QR^{-1}$ into the following form:
        $$P^T Q R^{-1} P = \left[\begin{array}{c:c} F_{a\times a} & * \\ \hdashline \mathbf{0} & A_{b\times b} \end{array}\right], P^T = \left[\begin{array}{c} * \\ \hdashline H_{b\times n} \end{array}\right] \quad (7)$$
        for an anti-stable matrix $F$ with an eigenvalue at the origin, stable matrix $A$, and $*$ denoting an arbitrary sub-matrix. The ordered real Schur form (available in Lapack, Matlab, and Octave software packages) is one alternative means of obtaining (7); see [30], [32] and the references therein for the Schur form and its numerical stability.
    Step 2: Solve for the $1 \times b$ vector $\boldsymbol{g}$ and $1 \times a$ vector $\boldsymbol{d}$ from the following linear matrix equation:
        $$\left[\begin{array}{c:c} \boldsymbol{g} & \boldsymbol{d} \end{array}\right] \left[\begin{array}{c:c} HR & -A^{-1}H\mathbf{1}_n \\ \hdashline -\tilde{Q}^* & \mathbf{1}_a \end{array}\right] = \left[\begin{array}{c:c} \mathbf{0}_{1\times n} & 1 \end{array}\right],$$
        with $\tilde{Q}^*$ denoting the matrix composed of the last $a$ rows of $\tilde{Q}$.
    Step 3: Write $\boldsymbol{c} = [\mathbf{0}, \boldsymbol{d}]$ and the steady-state joint pdf vector as
        $$\begin{aligned} \boldsymbol{f}(x) &= \boldsymbol{g}e^{Ax}Hu(x) + \boldsymbol{c}\delta(x), \\ f_i(x) &= \boldsymbol{g}e^{Ax}\boldsymbol{h}_i u(x) + c_i\delta(x), \end{aligned} \quad (8)$$
        where $\boldsymbol{h}_i$ denotes the $i^{\text{th}}$ column of $H$.
end function
```

the first two steps are $\mathcal{O}(n^3)$ in Alg. 1. In an important subcase that will be shown to arise in AoI queueing models later in this paper, a simple explicit way of complexity $\mathcal{O}(n^2)$ to find a matrix P in (7) of Alg. 1 is presented in the following lemma based on the Householder transformation given in [30].

Lemma 2: Consider the process $\boldsymbol{X}(t) \sim GMFQ(Q, \tilde{Q}, R)$ with order n, $R = \boldsymbol{diag}\{\boldsymbol{I}, -1\}$, and the last row of Q is the zero matrix. Let $\boldsymbol{u_1}$ be a column vector of ones of size n except for the last entry which is minus one. Also, let $\boldsymbol{u_2}$ be a column vector of zeros except for the first entry which is one. Let $\boldsymbol{u} = \boldsymbol{u_1} - ||\boldsymbol{u_1}||_2\boldsymbol{u_2}$. Then, the symmetric orthogonal matrix P defined by $P = \boldsymbol{I} - \frac{2\boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}}}{\boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{u}}$ gives rise to the factorization (7) *with the scalar F being zero.*

Other than the steady-state joint pdf and pma vectors, we are also interested in the following steady-state conditional density, denoted by $g_j^{\mathcal{S}_0}(x)$, of the fluid level just before a visit from any state in a particular subset $\mathcal{S}_0 \subset \mathcal{S}$ to one particular state $j \in \mathcal{S} \setminus \mathcal{S}_0$:

$$\frac{d}{dx} \lim_{\substack{t\to\infty \\ \Delta t\to 0}} \Pr\{X_f(t) \le x \mid X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\}, \tag{9}$$

for $x \ge 0$. The following theorem provides a closed-form expression to obtain the conditional steady-state density $g_j^{\mathcal{S}_0}(x)$ from the steady-state joint density $f_i(x)$ of the MFQ $\boldsymbol{X}(t)$.

Theorem 1: Let $\boldsymbol{X}(t) \sim GMFQ(Q, \tilde{Q}, R)$ be of order n and $\boldsymbol{f}(x)$ be its steady-state joint pdf vector given as in (8)*. Let $\mathcal{S}_0 \subset \mathcal{S}$ be such that the fluid level does not have a probability mass at zero in any of the states in $\mathcal{S}_0$. Then, the steady-state conditional density $g_j^{\mathcal{S}_0}(x)$ is written in terms of the steady-state density vector $\boldsymbol{f}(x)$:*

$$g_j^{\mathcal{S}_0}(x) = \frac{\sum\limits_{i\in\mathcal{S}_0} f_i(x)Q_{i,j}}{\int\limits_0^\infty \sum\limits_{i\in\mathcal{S}_0} f_i(x')Q_{i,j}dx'} = \boldsymbol{g'}e^{Ax}\boldsymbol{h}u(x), \tag{10}$$

where $\boldsymbol{h} = H\boldsymbol{\eta}$, $\eta_i = Q_{i,j}$ if $i \in \mathcal{S}_0$ and zero otherwise, and $\boldsymbol{g'} = \boldsymbol{g}/(-\boldsymbol{g}A^{-1}\boldsymbol{h})$.

Proof: We first write

$$\begin{aligned} &\lim_{\substack{t\to\infty \\ \Delta t\to 0}} \Pr\{X_f(t) \le x \mid X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\} \\ &= \lim_{\Delta t\to 0} \frac{\lim\limits_{t\to\infty} \Pr\{X_f(t) \le x, X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\}}{\lim\limits_{t\to\infty} \Pr\{X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\}}. \end{aligned} \tag{11}$$

Note that as Δt approaches 0, the denominator of above can be written as:

$$\begin{aligned} &= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\}\gamma_i(t), \\ &= \sum_{i\in\mathcal{S}_0} Q_{i,j}\Delta t \cdot \int_0^\infty f_i(x')dx' = \Delta t \int_0^\infty \sum_{i\in\mathcal{S}_0} f_i(x')Q_{i,j}dx', \end{aligned}$$

where $\gamma_i(t) = \Pr\{X_m(t) = i\}$. On the other hand, as Δt approaches 0, the numerator of the expression (11) is written as:

$$\begin{aligned} &\lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x, X_m(t+\Delta t) = j \mid X_m(t) = i\}\gamma_i(t) \\ &= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x \mid X_m(t) = i\}\zeta_{ij}(t)\gamma_i(t), \\ &= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \frac{\Pr\{X_f(t) \le x,\ X_m(t) = i\}}{\Pr\{X_m(t) = i\}}\zeta_{ij}(t)\gamma_i(t), \\ &= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x, X_m(t) = i\}\zeta_{ij}(t), \\ &= \Delta t \sum_{i\in\mathcal{S}_0} Q_{i,j} \int_0^x f_i(x')dx', \end{aligned}$$

where $\zeta_{ij}(t) = \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\}$. The first expression for $g_j^{\mathcal{S}_0}(x)$ in (10) immediately follows. The choice of $\boldsymbol{h}$ in the second expression (10) is that the numerator of the first expression in (10) can be written in the form $\boldsymbol{g}e^{Ax}\boldsymbol{h}u(x)$. The choice of $\boldsymbol{g'}$ is to ensure that the second expression is a legitimate density, i.e., the denominator of the first expression is $-\boldsymbol{g}A^{-1}\boldsymbol{h}$. ■

III. Queueing Models for the AoI and PAoI Processes

A. Bufferless Queues

In this subsection, we study the exact distributions of the AoI and PAoI processes in the $PH/PH/1/1/P(p)$ queue

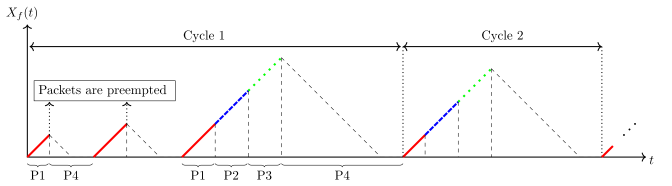

Fig. 3. Sample path of the fluid level process $X_f(t)$ where Pj denotes phase j, $j = 1, \dots, 4$.

which is formally described as follows. We assume the interrarrival time $\Lambda \sim PH(\boldsymbol{\tau}, T)$ with order k with $\boldsymbol{\kappa} = -T\mathbf{1}, \tau_0 = 1 - \boldsymbol{\tau}\mathbf{1} = 0$, and the service time $\Theta \sim PH(\boldsymbol{\sigma}, S)$ with order ℓ and $\boldsymbol{\nu} = -S\mathbf{1}, \sigma_0 = 1 - \boldsymbol{\sigma}\mathbf{1} = 0$. The information packet in service is to be preempted by a new information packet arrival with probability p, $0 \leq p \leq 1$.

For the purpose of obtaining the exact distributions of the AoI and PAoI processes in this system, we construct a GMFQ process $\boldsymbol{X}(t)$ by which we have a single fluid level trajectory of infinitely many cycles where each cycle begins with the arrival of an information packet into the system, proceeds until the reception of not the next but the second next successful information packet, and terminates with a final phase required for preparation for the next cycle. A sample path for the fluid level process $X_f(t)$ is depicted in Fig. 3 which is constructed as follows. At the beginning of phase 1 (solid red curve), an information packet, say packet i, arrives to an empty system and starts receiving service. During phase 1, this packet is in service and the fluid level $X_f(t)$ increases at a unit rate. When the packet i is not preempted during its service time (for example in Cycle 2 in Fig. 3), a transition to phase 2 occurs (dashed blue curve) which lasts until a new information packet arrival and in phase 2, the fluid level still increases at a unit rate. Upon a new arrival, a transition to phase 3 occurs (dotted green curve) during which the fluid level continues to increase at a unit rate until the service of the new packet arrival is over. If this packet turns out to be preempted during phase 3, the service time needs a reset. When phase 3 is over, the system transitions to phase 4 (dashed black curve) in which the fluid level is brought down to zero with a drift of minus one after which the fluid level stays at level 0 for an exponentially distributed duration of time with an arbitrary parameter (we use the unit parameter without loss of generality in this paper). When phase 4 is over, a cycle is completed and a new cycle gets to start again comprising phases one to four. On the other hand, if the first packet i in a cycle is to be preempted by a new packet, a direct transition into phase 4 occurs without experiencing the phases 2 and 3. This situation happens twice in Cycle 1 of Fig. 3. Irrespective of the type of such cycles, our main observation is the following. The value of the process $X_f(t)$ at the beginning of phase 2 is the system time D_j of a successful packet, say j. On the other hand, the value of $X_f(t)$ at the end of phase 3 (the reception epoch of packet $j+1$) is Φ_{j+1}. Therefore, the concatenation of the blue-dashed and green-dotted curves (from the beginning of phase 2 to the end of phase 3, with the phases 1 and 4 of each cycle excluded) in Fig. 3 is also a sample cycle of the AoI process in Fig. 2, and the fluid level just at the beginning of phase 4 in Fig. 3 is also a sample value of the PAoI process in Fig. 2, and vice versa. Moreover, the process $X_f(t)$ is the fluid level sub-process of a GMFQ process $\boldsymbol{X}(t) = (X_f(t), X_m(t))$ whose characterizing matrices we will now present. The state-space $\mathcal{S}$ of the modulating process $X_m(t)$ is written as $\mathcal{S} = \bigcup_{n=1}^{4} \mathcal{S}_n$ where the set $\mathcal{S}_n$ refers to the set of states to be used for phase n and is given as follows:

$$\begin{aligned} \mathcal{S}_m &= \{(i^{(m)}, j^{(m)})\}, \quad i^{(m)} = 1, \dots, k, \ j^{(m)} = 1, \dots, \ell, \\ \mathcal{S}_2 &= \{i^{(2)}\}, \quad i^{(2)} = 1, \dots, k, \end{aligned} \tag{12}$$

for $m = 1, 3$ and $\mathcal{S}_4 = \{0\}$. For the set $\mathcal{S}_m$, $i^{(m)}$ ($j^{(m)}$) keeps track of the state of the interarrival time (service time). With the lexicographical ordering of the states from $\mathcal{S}_1$ to $\mathcal{S}_4$, $\boldsymbol{X}(t) \sim GMFQ(Q, \tilde{Q}, R)$ of order $2k\ell + k + 1$ where

$$\begin{aligned} Q &= \left[\begin{array}{c:c:c:c} Q_{11} & \boldsymbol{I}_k \otimes \boldsymbol{\nu} & \mathbf{0} & p\boldsymbol{\kappa} \otimes \mathbf{1}_\ell \\ \hdashline \mathbf{0} & T & \boldsymbol{\kappa} \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma}) & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & Q_{33} & \mathbf{1}_k \otimes \boldsymbol{\nu} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{array}\right], \\ R &= \mathbf{diag}\{\boldsymbol{I}_{2k\ell+k}, -1\}, \end{aligned} \tag{13}$$

where

$$\begin{aligned} Q_{11} &= \boldsymbol{I}_k \otimes S + T \otimes \boldsymbol{I}_\ell + (1-p)(\boldsymbol{\kappa} \otimes \boldsymbol{\tau}) \otimes \boldsymbol{I}_\ell, \\ Q_{33} &= Q_{11} + p\boldsymbol{\kappa} \otimes (\mathbf{1}_\ell \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma})). \end{aligned} \tag{14}$$

This characterization is based on the verbal description of each of the four phases according to Fig. 3. To see this, let us assume that we are in a phase 1 state $(i^{(1)}, j^{(1)})$. A new arrival occurring with transition rate $\boldsymbol{\kappa}_{i^{(1)}}$ will lead us (i) to state 0 with probability p which explains the term $p\boldsymbol{\kappa} \otimes \mathbf{1}_\ell$ (ii) to state $(v, j^{(1)})$ with probability $(1-p)\tau_v$ which explains the term $(1-p)(\boldsymbol{\kappa} \otimes \boldsymbol{\tau}) \otimes \mathbf{I}_\ell$. When the service time is over which occurs with transition rate $\boldsymbol{\nu}_{j^{(1)}}$, we transition to a phase 2 state $i^{(2)}$ which is the same as $i^{(1)}$ which explains the term $\boldsymbol{I}_k \otimes \boldsymbol{\nu}$. Other state transitions without new arrivals or departures are reflected in Q_{11}. One can follow this procedure to justify the other blocks of the matrix Q. The expression for R immediately follows the verbal description of the four phases in the sample paths. $\tilde{Q}$ is a matrix of zeros except for the $1 \times k\ell$ block at the south-west corner of Q which should be set to $\boldsymbol{\tau} \otimes \boldsymbol{\sigma}$ to start a new cycle and the scalar at the south-east corner should be set to minus one. To see this, we escape from state 0 to state $(i^{(1)}, j^{(1)})$ with rate $\boldsymbol{\tau}_{i^{(1)}} \boldsymbol{\sigma}_{j^{(1)}}$ which explains the term $\boldsymbol{\tau} \otimes \boldsymbol{\sigma}$ and then the row sum of the last row is set to zero by this choice of the south-east corner. Due to the sample cycle arguments described above, the steady-state solution of $\boldsymbol{X}(t)$ enables one to obtain the exact distributions of the AoI and PAoI processes as stated in the next theorem.

Theorem 2: Consider the $GMFQ(Q, \tilde{Q}, R)$ with the characterizing matrices as defined in (13) *with $f_s(x), s \in \mathcal{S}$ being the steady-state joint pdf for the $GMFQ(Q, \tilde{Q}, R)$ and can be found using Alg. 1 and Lemma 2 stemming from the structure of Q and R. Then, the AoI and PAoI processes are ME-distributed and their pdfs, denoted by $f_\Delta(x)$ and $f_\Phi(x)$, respectively, for the $PH/PH/1/1/P(p)$ queue, are given in* *terms of $f_s(x)$ as follows:*

$$f_\Delta(x) = \frac{\sum\limits_{s \in \mathcal{S}_2 \bigcup \mathcal{S}_3} f_s(x)}{\int\limits_0^\infty \sum\limits_{s \in \mathcal{S}_2 \bigcup \mathcal{S}_3} f_s(x')\,dx'},$$

$$f_\Phi(x) = \frac{\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{\ell} f_{(i^{(3)},j^{(3)})}(x)\nu_j}{\int\limits_0^\infty \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{\ell} f_{(i^{(3)},j^{(3)})}(x')\nu_j\,dx'}. \tag{15}$$

Proof: The proof follows sample path arguments. The expression for $f_\Delta(x)$ stems from the observation that the restriction of the $GMFQ(Q,\tilde{Q},R)$ to the states in $\mathcal{S}_2$ and $\mathcal{S}_3$, as indicated by the blue-dashed and green-dotted parts of the curve in Fig. 3, comprises the cycles of the actual AoI process. On the other hand, $f_\Phi(x)$ amounts to the pdf of the fluid level just at the epoch of transitions to state 0 from the states in $\mathcal{S}_3$ and is therefore given by the expression in (15) using Theorem 1. The ME-distributed nature of the stationary AoI and PAoI is an immediate result of Step 3 of Alg. 1 as well as Theorem 1. ■

B. Single Buffer Queues

In this subsection, we study the $M/PH/1/2/R(r)$ queue model without preemption, and with a waiting room for one single packet only. An information packet joins the queue upon arrival if the buffer is empty but the packet in the buffer is to be replaced with probability r, $0 \leq r \leq 1$, by the new incoming information packet arrival. We assume the interarrival time is exponentially distributed with parameter λ and the service time $\Theta \sim PH(\boldsymbol{\sigma}, S)$ with order ℓ and $\boldsymbol{\nu} = -S\mathbf{1}, \sigma_0 = 1 - \boldsymbol{\sigma}\mathbf{1} = 0$. In order to derive the pdf of the queue wait time W for successful packets, denoted by $f_W(x), x \geq 0$, we introduce a new fluid level process $Y_f(t)$ a sample path of which is given in Fig. 4 that is related to the residual service time. For this purpose, let us assume a packet arrival at $t = 0$ to an empty queue. Upon the arrival, the fluid process starts to increase at a unit rate for a duration that equals to the service time. This behavior is said to occur in phase 1 (shown by the dashed blue curve). Once the service time is over, a transition to phase 2 (shown by the dotted green curve) occurs in which the fluid level is allowed to decrease with a rate of minus one. If a new arrival occurs in phase 2 when $Y_f(t) > 0$, we transition to phase 3 (shown by the solid red curve) during which $Y_f(t)$ is reduced at a unit rate irrespective of new arrivals. If we hit zero without transitioning to phase 3, with a new arrival, a transition from phase 2 to phase 1 occurs. When we hit zero at phase 3, we stay at level 0 for an exponentially distributed duration of time with unit parameter after which we transition to phase 1. If we exclude phase 1 and part of phase 3 corresponding to $Y_f(t) = 0$ in this sample path, what remains is a sample path for the residual service time, i.e., the time needed to drain the packet in service, for both queues. The residual service time is zero if there is no packet in service.

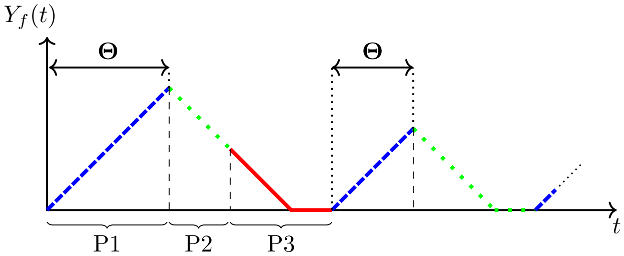

Fig. 4. Sample path of the fluid level process $Y_f(t)$ corresponding to the residual service time. Pj denotes phase j, $j = 1, \ldots, 3$.

Actually, $Y_f(t)$ is the fluid level subprocess of a GMFQ process $\boldsymbol{Y}(t) = (Y_f(t), Y_m(t))$ where $Y_m(t) \in \mathcal{S} = \bigcup_{n=1}^{3} \mathcal{S}_n$ where $\mathcal{S}_1 = \{1^{(1)}, \ldots, \ell^{(1)}\}$, $\mathcal{S}_2 = \{0\}$, and $\mathcal{S}_3 = \{1\}$. With the enumeration of the states from $\mathcal{S}_1$ to $\mathcal{S}_3$ as before, the proposed GMFQ is characterized with the triple $(Q, \tilde{Q}, R)$ of order $\ell + 2$ where

$$Q = \left[\begin{array}{c:c:c} S & \boldsymbol{\nu} & \mathbf{0} \\ \hdashline \mathbf{0} & -\lambda & \lambda \\ \hdashline \mathbf{0} & 0 & 0 \end{array}\right], \quad \tilde{Q} = \left[\begin{array}{c:c:c} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \lambda\boldsymbol{\sigma} & -\lambda & 0 \\ \hdashline \boldsymbol{\sigma} & 0 & -1 \end{array}\right], \tag{16}$$

and $R = \mathbf{diag}\{\boldsymbol{I}_l, -\boldsymbol{I}_2\}$. Let $f_s(x), s \in \mathcal{S}$ denote the steady-state joint density for the $GMFQ(Q, \tilde{Q}, R)$ obtained using the method described in subsection II-C. Based on Step 3 of Alg. 1, we have the following matrix-exponential form for $f_i(x)$:

$$f_i(x) = \boldsymbol{g}e^{Ax}\boldsymbol{h}_i u(x) + c_i\delta(x), \quad i = 0, 1, \tag{17}$$

for a row vector $\boldsymbol{g}$ of size ℓ, a square matrix A of size ℓ, a column vector $\boldsymbol{h}_i$ of size ℓ, and two scalars c_0 and c_1 corresponding to the probability masses at states 0 and 1, respectively. The following theorem gives an expression for $f_W(x)$ for the two queueing systems of interest based on the expression (17).

Theorem 3: The pdf of the queue wait time $f_W(x)$ for the $M/PH/1/2/R(r)$ queue is given in terms of the matrix parameters of (17) *as follows:*

$$f_W(x) = \eta_1 \left(\boldsymbol{g}e^{(A - r\lambda\boldsymbol{I}_\ell)x}(\boldsymbol{h}_0 + r\boldsymbol{h}_1)u(x) + c_0\delta(x)\right), \tag{18}$$

where $\eta_1 = 1/\left(-\boldsymbol{g}(A - r\lambda\boldsymbol{I}_\ell)^{-1}(\boldsymbol{h}_0 + r\boldsymbol{h}_1) + c_0\right)$.

Proof: Let $\tilde{f}_i(x)$, $i = 0, 1$ denote the joint pdf of the residual service time and there are i information packets in the queue. We censor out the states in $\mathcal{S}_1$ and the probability mass at zero for state 1 from the original GMFQ to obtain the following expression for $\tilde{f}_i(x)$:

$$\begin{aligned} \tilde{f}_0(x) &= \eta\left(\boldsymbol{g}e^{Ax}\boldsymbol{h}_0 u(x) + c_0\delta(x)\right), \\ \tilde{f}_1(x) &= \eta\boldsymbol{g}e^{Ax}\boldsymbol{h}_1 u(x), \ x \geq 0, \end{aligned}$$

where

$$\eta = 1/\left(-\boldsymbol{g}A^{-1}(\boldsymbol{h}_0 + \boldsymbol{h}_1) + c_0\right).$$

For the $M/PH/1/2/R(r)$ queue, a successful packet arrival occurs when the following conditions hold: (i) there are no information packets in the queue or existing packet is replaced with this new arrival, and (ii) this packet arriving when the residual service time equals $x \geq 0$, will not be replaced

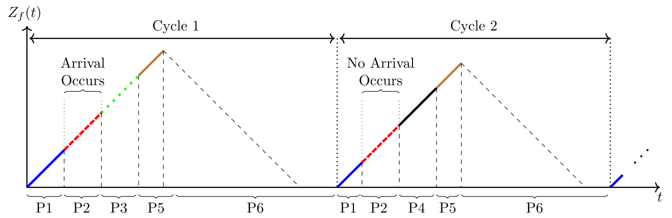

Fig. 5. Sample path of the fluid level process $Z_f(t)$ where Pj denotes phase j, $j = 1, \ldots, 6$.

with another packet which occurs with probability $e^{-r\lambda x}$. Therefore, for the $M/PH/1/2/R(r)$ queue,

$$f_W(x) = \frac{e^{-r\lambda x}}{\pi_s}\left(\tilde{f}_0(x) + r\tilde{f}_1(x)\right), \quad x \geq 0,$$

where $\pi_s = \int_{-\infty}^{\infty}(\tilde{f}_0(x) + r\tilde{f}_1(x))e^{-r\lambda x}dx$ is the success probability which consequently yields the expression (18) completing the proof. ■

Remark 1: The scalar term $e^{-r\lambda x}$ in the above proof arising when arrivals are Poisson, commutes with the matrix e^{Ax} giving rise to the simple expression in Theorem 3. This is the main reason that hindered us from using more general PH-type interarrivals for single-buffer systems in which case such commutativity would not hold.

The pdf given in the identity (18) is in matrix exponential form. Therefore, W can be shown to be ME-distributed and $W \sim ME(\boldsymbol{\beta}^*, B^*)$ of order ℓ for some vector $\boldsymbol{\beta}^*$ and matrix B^* using the construction method described in Lemma 1. Let us first assume $W \sim PH(\boldsymbol{\beta}, B)$ with $\boldsymbol{\psi} = -B\mathbf{1}$, $\beta_0 = 1 - \boldsymbol{\beta}\mathbf{1}$ with order ℓ. Note that computational techniques and algorithms using PH-type distributions can be extended to ME distributions as well, as shown in various studies; see [28] and the references therein. Therefore, after developing a numerical algorithm for obtaining $f_\Delta(x)$ and $f_\Phi(x)$ under the assumption $W \sim PH(\boldsymbol{\beta}, B)$, we will then substitute $(\boldsymbol{\beta}^*, B^*)$ in place of $(\boldsymbol{\beta}, B)$ in the numerical algorithm when $W \sim ME(\boldsymbol{\beta}^*, B^*)$.

A sample path of the fluid level process $Z_f(t)$ we propose to use for the $M/PH/1/2/R(r)$ queue is depicted in Fig. 5. Similar to the $PH/PH/1/1/P(p)$ case, we have a single fluid level trajectory of infinitely many cycles where each cycle begins with the arrival of a successful information packet into the system and ends with the reception of the next successful information packet at the destination. Let us assume that at $t = 0$, a successful information packet arrival has just occured. In phase 1 (shown by the solid blue curve), the fluid level $Z_f(t)$ increases at a unit rate until its queue wait time W is over. Once phase 1 is over, phase 2 starts (shown by the dashed red curve) which lasts until either a new packet arrival occurs or until the service time of the current packet is over. In the former case, we transition into phase 3 (shown by the solid black curve) during which we wait for the end of the service time before transitioning to phase 4. In the latter case, we transition from phase 2 to directly phase 4 (shown by the dotted green curve) in which case it lasts until a new packet arrival. In phase 5 (shown by the solid brown curve), the fluid level continues to increase at a unit rate until the service time of the new packet arrival is over. When phase 5 terminates, the system transitions to phase 6 in which the fluid level is brought down to zero with a drift of minus one after which the fluid level stays at level 0 for an exponentially distributed duration of time with unit parameter. Similar to the sample path observations made for the bufferless system, the parts of the curves from the beginning of phase 4 until the end of phase 5 in each cycle in Fig. 5 is also a sample cycle of the AoI process of Fig. 2 for the single buffer queueing system, and the fluid level just at the beginning of phase 6 in Fig. 5 is also a sample value of the PAoI process of Fig. 2, and vice versa. Actually, $Z_f(t)$ is the fluid level sub-process of a GFMQ $\boldsymbol{Z}(t) = (Z_f(t), Z_m(t))$ where $Z_m(t) \in \mathcal{S} = \bigcup_{n=1}^{6} \mathcal{S}_n$ where the set $\mathcal{S}_n$ refers to the collection of states to be used for phase n. In particular, $\mathcal{S}_m = \{i^{(m)}\}$, $i^{(m)} = 1, \ldots, \ell$, $m = 1, 2, 3, 5$, $\mathcal{S}_4 = \{1\}$, and $\mathcal{S}_6 = \{0\}$. With the enumeration of the states from $\mathcal{S}_1$ to $\mathcal{S}_6$ as before, the proposed GMFQ is characterized with the triple $(Q, \tilde{Q}, R)$ of order $4\ell + 2$ where

$$Q = \left[\begin{array}{c:c:c:c:c:c} B & \boldsymbol{\psi} \otimes \boldsymbol{\sigma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & S - \lambda \boldsymbol{I}_\ell & \lambda \boldsymbol{I}_\ell & \boldsymbol{\nu} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & S & \mathbf{0} & \boldsymbol{\nu} \otimes \boldsymbol{\sigma} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & -\lambda & \lambda \boldsymbol{\sigma} & 0 \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & S & \boldsymbol{\nu} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} & 0 \end{array}\right], \tag{19}$$

$R = \textbf{diag}\{\boldsymbol{I}_{4\ell+1}, -1\}$, and $\tilde{Q}$ is a matrix of zeros except for the $1 \times 2\ell$ block at the south-west corner which should be set to $[\boldsymbol{\beta}, \beta_0 \boldsymbol{\sigma}]$ and the scalar at the south-east corner should be -1. Let $f_s(x), s \in \mathcal{S}$ denote the steady-state joint density for the $GMFQ(Q, \tilde{Q}, R)$. The expressions for the pdfs of the AoI and PAoI processes are given in the following theorem, a proof of which is similar to that of Theorem 2 and is omitted.

Theorem 4: Consider the process $\boldsymbol{Z(t)} \sim GMFQ$ $(Q, \tilde{Q}, R)$ with the characterizing matrices as defined in (19) with $f_s(x), s \in \mathcal{S}$ denoting the associated steady-state joint pdf. The pdfs of the steady-state AoI and PAoI, denoted by $f_\Delta(x)$ and $f_\Phi(x)$, respectively, for the $M/PH/1/2/R(r)$ queue, are in matrix exponential form and are given in terms of $f_s(x)$ as follows:

$$f_\Delta(x) = \frac{\sum\limits_{s \in \mathcal{S}_4 \bigcup \mathcal{S}_5} f_s(x)}{\int\limits_0^\infty \sum\limits_{s \in \mathcal{S}_4 \bigcup \mathcal{S}_5} f_s(x')dx'},$$

$$f_\Phi(x) = \frac{\sum\limits_{j=1}^{\ell} f_{j^{(5)}}(x)\nu_j}{\int\limits_0^\infty \sum\limits_{j=1}^{\ell} f_{j^{(5)}}(x')\nu_j dx'}. \tag{20}$$

IV. Numerical Examples

In order to make the numerical experiments reproducible, we made our Matlab implementation available at MATLAB File Exchange at URL https://www.mathworks.com/matlabcentral/fileexchange/74923-queuing-models-of-age-of-information with the title "Queueing Models of Age of Information".

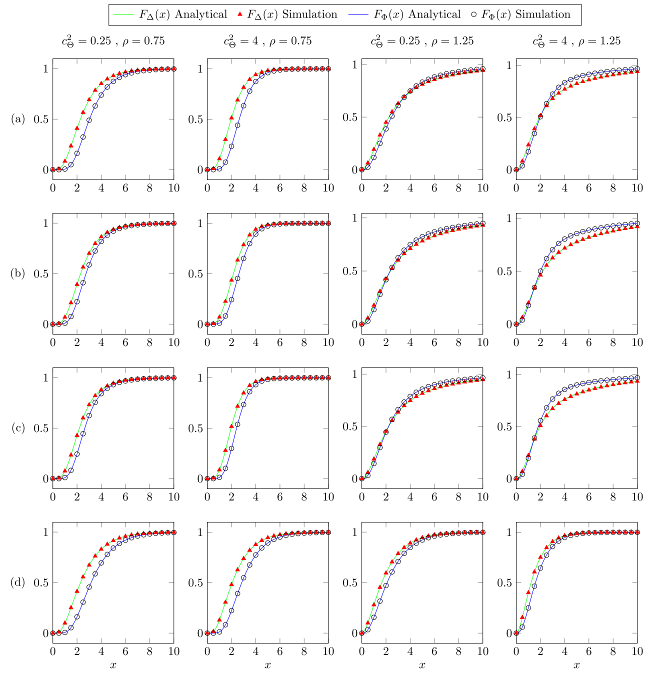

Fig. 6. The cdfs of the AoI and PAoI processes obtained with the proposed method and simulations for the scenarios $\rho = 0.75, 1.25$, $c^2_\Theta = 0.25, 4$ for (a) $M/PH/1/1$ (b) $M/PH/1/1^*$ (c) $M/PH/1/2$ (d) $M/PH/1/2^*$ queueing models.

A. Validation With Simulations

In the first example, we assume Poisson arrivals with rate λ and fix $p = 0$ or $p = 1$ for the bufferless system and $r = 0$ or $r = 1$ for the single-buffer system, and then obtain the steady-state distributions of the AoI and the PAoI of the underlying four queueing systems, namely $M/PH/1/1$, $M/PH/1/1^*$, $M/PH/1/2$, and $M/PH/1/2^*$ queues, and compare with simulations. For the service times, we use a PH-type distribution with mean ρ/λ for a given system load ρ and scov of the service time is to be fixed to a given value c^2_Θ according the following procedure: For $c^2_\Theta = 1/j \leq 1$ for a positive integer j, $E(\mu^{-1}, j)$ distribution is used which refers to an Erlang distribution with mean μ^{-1} and with order j. If j is not an integer, then we resort to a mixture of two appropriate Erlang distributions [33]. When $c^2_\Theta > 1$, then we propose to use a hyper-exponential distribution with balanced means to fit the first two moments [33]. Fig. 6 depicts the cdfs of the AoI and PAoI processes obtained with the proposed analytical model as well as simulations, for four different values of the parameter pair (ρ, c^2_Θ), and for the four queueing models and perfect match with the simulation results is obtained in all cases.

As a second example, we allow PH-type information packet arrivals for the bufferless case and the same construction of the arrival process is used as in the previous example to match λ and c^2_Λ. The cdfs of the AoI and PAoI processes are plotted in Fig. 7 using the proposed analytical model and simulations for $\rho = 0.75, 1.25$, $c^2_\Lambda = 0.25, 4$ while fixing $c^2_\Theta = 0.2$ for the two queueing models $PH/PH/1/1$ and $PH/PH/1/1^*$. The results obtained by the analytical method perfectly match those obtained by simulations.

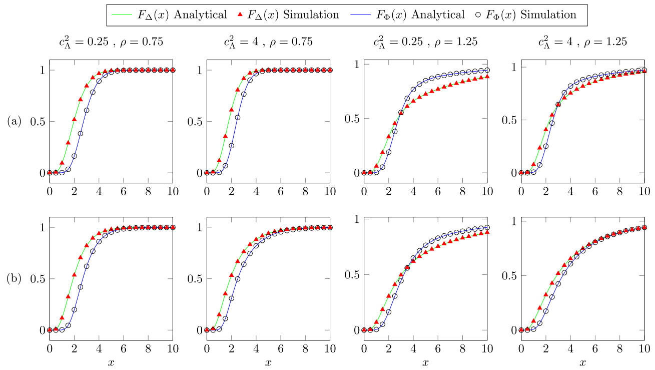

Fig. 7. The cdfs of the AoI and PAoI processes when $c^2_\Theta = 0.2$, $\rho = 0.75, 1.25$, $c^2_\Lambda = 0.25, 4$ for the (a) $PH/PH/1/1$ (b) $PH/PH/1/1^*$ queueing models.

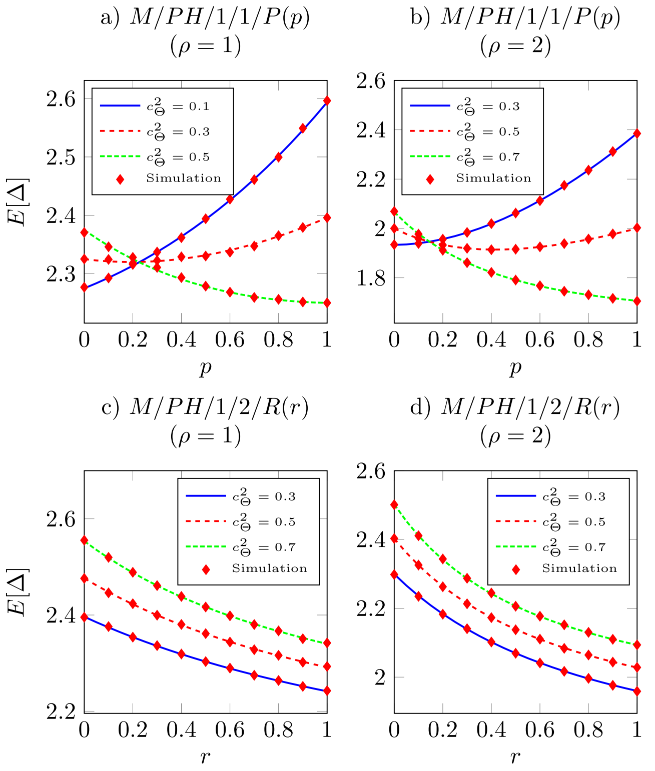

Fig. 8. The mean AoI, as a function of the packet preemption probability p for the $M/PH/1/1/P(p)$ queue for three values of c^2_Θ: (a) $\rho = 1$ (b) $\rho = 2$, and as a function of the packet replacement probability for the $M/PH/1/2/R(r)$ queue again for three values of c^2_Θ: (c) $\rho = 1$ (d) $\rho = 2$.

As a final example of this subsection, the mean AoI, $E[\Delta]$, is plotted in Fig. 8 (obtained with both the proposed method and simulations which perfectly match) with respect to the packet preemption probability p for the $M/PH/1/1/P(p)$ queue and with respect to the packet replacement probability r for the $M/PH/1/2/R(r)$ queue, each queue with varying values of ρ and c^2_Θ. From the examples we studied, we observe that there is an optimal choice of the preemption probability (not necessarily 0 or 1) which minimizes $E[\Delta]$ and this optimal value p^* depends on the pair (ρ, c^2_Θ). As a general remark, the complete preemption policy $p = 1$ turns out to

TABLE I

THE PERFORMANCE METRICS $E[\Delta]$ AND $E[\Delta^2]$ OBTAINED WITH THE CLOSED-FORM EXPRESSIONS IN REF. [4] AND THE PROPOSED METHOD FOR VARIOUS QUEUEING MODELS AND THEIR PARAMETERS

		$E[\Delta]$	
Queueing Model		Ref. [4]	Prop.
$M/PH/1/1^*$	$\lambda = 0.5,\ \Theta \sim E(1,2)$	3.1250	3.1250
	$\lambda = 0.5,\ \Theta \sim E(1,4)$	3.2036	3.2036
	$\lambda = 1.5,\ \Theta \sim E(1,2)$	2.0417	2.0417
	$\lambda = 1.5,\ \Theta \sim E(1,4)$	2.3830	2.3830
$PH/M/1/1^*$	$\mu = 0.5,\ \Lambda \sim E(1,2)$	2.7500	2.7500
	$\mu = 1.5,\ \Lambda \sim E(1,2)$	1.4167	1.4167
	$\mu = 0.5,\ \Lambda \sim E(1,4)$	2.6250	2.6250
	$\mu = 1.5,\ \Lambda \sim E(1,4)$	1.2917	1.2917
$M/PH/1/2^*$	$\lambda = 0.5,\ \Theta \sim E(1,2)$	3.1089	3.1089
	$\lambda = 0.5,\ \Theta \sim E(1,4)$	3.0786	3.0786
	$\lambda = 1.5,\ \Theta \sim E(1,2)$	2.0996	2.0996
	$\lambda = 1.5,\ \Theta \sim E(1,4)$	2.0226	2.0226
		$E[\Delta^2]$	
Queueing Model		Ref. [4]	Prop.
$M/PH/1/1^*$	$\lambda = 0.5,\ \Theta \sim E(1,2)$	14.5312	14.5312
	$\lambda = 0.5,\ \Theta \sim E(1,4)$	14.8310	14.8310
	$\lambda = 1.5,\ \Theta \sim E(1,2)$	6.0035	6.0035
	$\lambda = 1.5,\ \Theta \sim E(1,4)$	7.8910	7.8910
$PH/M/1/1^*$	$\mu = 0.5,\ \Lambda \sim E(1,2)$	12.0000	12.0000
	$\mu = 1.5,\ \Lambda \sim E(1,2)$	2.8889	2.8889
	$\mu = 0.5,\ \Lambda \sim E(1,4)$	11.1250	11.1250
	$\mu = 1.5,\ \Lambda \sim E(1,4)$	2.3472	2.3472

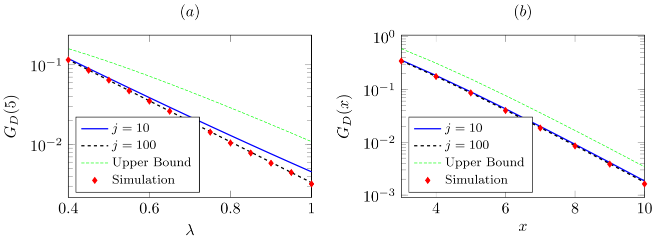

Fig. 9. The age violation probability $G_\Delta(x)$ (a) with respect to the arrival rate λ when $x = 5,\ \mu = 1$ and (b) with respect to the age limit x when $\lambda = 0.45,\ \mu = 1$, obtained by simulations and the proposed method along with upper bound of [20].

perform relatively poorly for larger values of ρ and smaller values of c^2_Θ. On the other hand, as expected, the complete replacement policy $r = 1$ gives rise to the minimum $E[\Delta]$ for all the examples we studied.

B. Validation With Existing Results

In addition to comparisons with simulations, we validate the accuracy of our proposed method with a number of closed-form expressions existing in the literature. As an example, closed-form expressions for $E[\Delta]$ and $E[\Delta^2]$ are available in [4] for the $M/GI/1/1^*$ and $GI/M/1/1^*$ queues as well as an expression for $E[\Delta]$ for the $M/GI/1/2^*$ queue. In Table I, we tabulate the results up to four fractional digits, obtained with our proposed method and by the closed-form expressions given in [4] for various cases. The results perfectly match in all digits and for all the cases we tried, demonstrating the accuracy of the proposed method.

As a second example, we compare in Fig. 9 our proposed analytical method for the $PH/D/1/1$ queue with $\Lambda \sim E(\lambda^{-1}, 2)$ against the upper bound proposed by [20] for the age violation probability $G_\Delta(x) = 1 - F_\Delta(x),\ x \geq 0$. For our method to work, we approximate the deterministic service time by the Erlang distribution $E(\mu^{-1}, j)$ with two relatively large values for the order parameter $j = 10, 100$. We use the same numerical example of [20] for which in Fig. 9(a), $G_D(5)$ is plotted with respect λ when $\mu = 1$ and in Fig. 9(b), $G_D(x)$ is plotted with respect to the age limit x when $\lambda = 0.45$ and $\mu = 1$. Fig. 9 demonstrates that the Erlang approximation with order $j = 100$ employed along with the proposed method is quite successful, whereas the easy to obtain upper bound is quite loose for this particular example as shown also in [20].

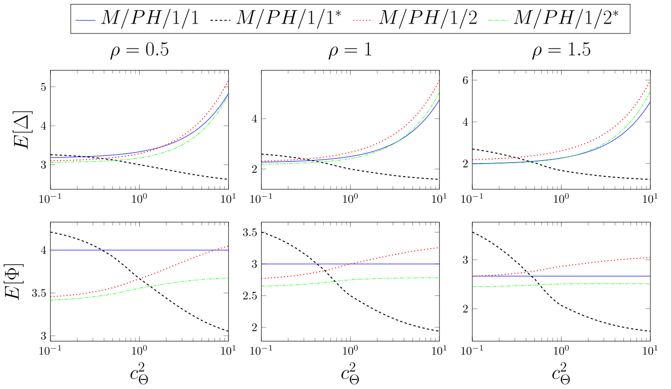

Fig. 10. The mean AoI and the mean PAoI as a function of the scov of the service time for the four proposed queueing models for three different values of $\rho = 0.5, 1, 1.5$.

C. Analytical Results

In this subsection, only results obtained with the proposed method are reported for the four queueing systems, namely $PH/PH/1/1$, $PH/PH/1/1^*$, $M/PH/1/2$, $M/PH/1/2^*$ queues, corresponding to the choices of the preemption parameter $p = 0$, $p = 1$ and replacement parameter $r = 0$, $r = 1$, respectively. As the first example of this subsection, the mean AoI and the mean PAoI are depicted in Fig. 10 as a function of the scov of the service time for the four queueing models of interest for three different values of the system load, namely $\rho = 0.5, 1, 1.5$, representative of light, critical, and high system load scenarios, respectively. We have the following observations:

- The $M/PH/1/1^*$ model outperforms all the other models in terms of minimizing both $E[\Delta]$ and $E[\Phi]$ for larger values of c^2_Θ. However, the $M/PH/1/1^*$ model performs quite poorly for smaller values of c^2_Θ, a situation which is emphasized even more, for larger system loads. This is not surprising since when $c^2_\Theta < 1$, the expected residual service time of the ongoing service time will tend to be smaller than $E[\Theta]$ and the benefit of preemption is to diminish in such situations.
- The PAoI metric $E[\Phi]$ in the $M/PH/1/1$ model exhibits insensitivity to the scov of the service time which is resemblant of the blocking probability in an $M/G/c/c$ system which is dependent on the service time only through its mean and not its higher moments.

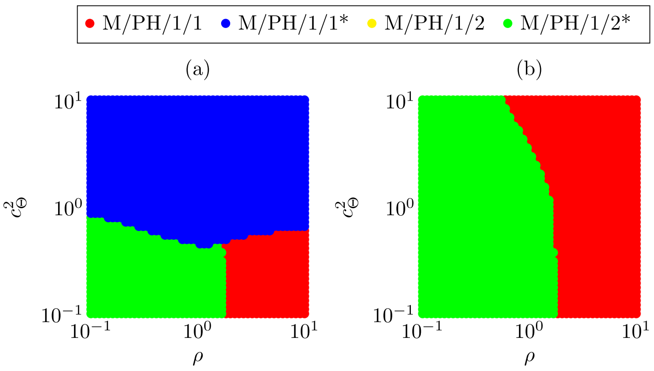

Fig. 11. The queueing model which minimizes the mean AoI as a function of the system load ρ and scov of the service time c_Θ^2 (a) out of all four queueing models (b) out of three queueing models when $M/PH/1/1^*$ is excluded.

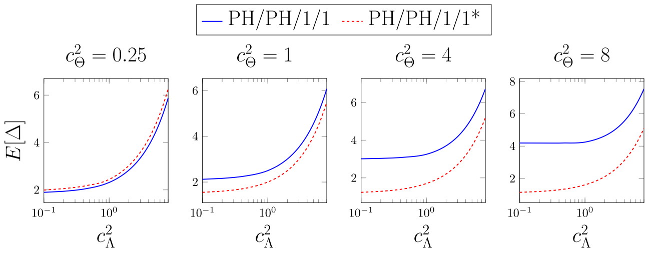

Fig. 12. Mean AoI for the $PH/PH/1/1$ and $PH/PH/1/1^*$ queues with respect to varying c_Λ^2 for four different values of c_Θ^2 when $\rho = 1$.

However, this feature is not inherited in the mean AoI which appears to increase with increased c_Θ^2.

- Both the AoI and PAoI metrics of interest increase with increased values of c_Θ^2 for the single-buffer $M/PH/1/2$ and $M/PH/1/2^*$ queues for three values of the system load that we studied, whereas the $M/PH/1/2^*$ system presents consistently better performance since a new information packet is always timelier at the remote server than the one already waiting in the queue.

In the second example, the queueing model (out of four) which gives rise to the minimum mean AoI is depicted in Fig. 11(a) as a function of the system load ρ and scov of the service time c_Θ^2. For larger values of c_Θ^2, the best system is the $M/PH/1/1^*$ model irrespective of the system load. However, for smaller values of c_Θ^2, the $M/PH/1/1$ model gives the best performance for larger values of the system load, whereas it is taken down by the $M/PH/1/2^*$ model for lower values of the system load. In some cases, preemption may not be possible since the information packet in service may not be under the control of the server once the service begins. Consequently, we depict the best queueing model out of three models only (when the preemptive $M/PH/1/1^*$ model is excluded) in Fig. 11(b) which shows that the boundary between the queueing models $M/PH/1/2^*$ and $M/PH/1/1$ turn out to depend on the particular value of c_Θ^2 when $c_\Theta^2 > 1$. The $M/PH/1/2$ model does not give rise to the best mean AoI figure in any of these plots since it is always outperformed by $M/PH/1/2^*$. As a final example, we study the impact of the scov of the interarrival times in the mean AoI. Fig. 12 depicts the mean AoI with respect to varying c_Λ^2 for four different values of c_Θ^2 for a critically loaded system with $\rho = 1$. Both queueing models tend to be affected adversely with increased c_Λ^2, but for larger values of c_Λ^2, the $PH/PH/1/1$ system is penalized more severely.

V. Conclusion

In this paper, we propose and validate a computationally efficient and stable numerical method for obtaining the exact steady-state distributions of both the AoI and PAoI processes for the bufferless $PH/PH/1/1/P(p)$ queue with probabilistic preemption and the $M/PH/1/2/R(r)$ queue with probabilistic replacement. Age violation probabilities and the associated moments can be calculated easily since the obtained distributions are in matrix exponential form. Cases for which probabilistic preemption is beneficial for bufferless systems with respect to complete preemption and no preemption cases are identified. For single buffer systems, the complete replacement policy is shown to yield the best AoI performance. The scov of the service time is shown to play an important role for all the queueing systems of interest on the mean values of AoI and PAoI with the exception that the mean PAoI for the $M/PH/1/1$ system appears to be insensitive to higher order moments of the service time. Future work will include extensions of the proposed model to update systems involving multiple sources.

References

[1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.

[2] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 326–330.

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

[4] 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 Trans. Inf. Theory*, vol. 65, no. 12, pp. 8305–8324, Dec. 2019.

[5] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Jun. 2011, pp. 350–358.

[6] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. 46th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2012, pp. 1–6.

[7] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.

[8] K. Chen and L. Huang, "Age-of-information in the presence of error," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2579–2583.

[9] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1681–1685.

[10] A. Arafa and S. Ulukus, "Age minimization in energy harvesting communications: Energy-controlled delays," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 1801–1805.

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

[12] Y.-P. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 561–565.

[13] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, Jul. 2018.

[14] R. D. Yates and S. Kaul, “Real-time status updating: Multiple sources,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2666–2670.

[15] A. M. Bedewy, Y. Sun, and N. B. Shroff, “Optimizing data freshness, throughput, and delay in multi-server information-update systems,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2569–2573.

[16] V. G. Kulkarni, “Fluid models for single buffer systems,” in *Frontiers in Queueing: Models and Applications in Science and Engineering*, J. H. Dshalalow, Ed. Boca Raton, FL, USA: CRC Press, 1997, pp. 321–338.

[17] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, “The stationary distribution of the age of information in FCFS single-server queues,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 571–575.

[18] E. Najm and R. Nasser, “Age of information: The gamma awakening,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2574–2578.

[19] A. Soysal and S. Ulukus, “Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization,” 2019, *arXiv:1905.13743*. [Online]. Available: https://arxiv.org/abs/1905.13743

[20] 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-IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 37–45.

[21] M. F. Neuts, *Matrix-Geometric Solutions Stochastic Models: An Algorithmic Approach*. New York, NY, USA: Dover, 1981.

[22] C. A. O’Cinneide, “Characterization of phase-type distributions,” *Commun. Statistics. Stochastic Models*, vol. 6, no. 1, pp. 1–57, Jan. 1990.

[23] S. Asmussen, O. Nerman, and M. Olsson, “Fitting phase-type distributions via the EM algorithm,” *Scandin. J. Statist.*, vol. 23, no. 4, pp. 419–441, 1996.

[24] D. Anick, D. Mitra, and M. M. Sondhi, “Stochastic theory of a data-handling system with multiple sources,” *Bell Syst. Tech. J.*, vol. 61, no. 8, pp. 1871–1894, Oct. 1982.

[25] L. Kosten, “Stochastic theory of data-handling systems with groups of multiple sources,” in *Performance of Computer-Communication Systems*, H. Ruding and W. Bux, Eds. Amsterdam, The Netherlands: Elsevier, pp. 321–331.

[26] S. Asmussen and M. Bladt, “Renewal theory and queueing algorithms for matrix-exponential distributions,” in *Matrix-Analytic Methods in Stochastic Models*, A. Alfa and S. Chakravarthy, Eds. New York, NY, USA: Marcel Dekker, 1996, pp. 313–341.

[27] Q.-M. He and H. Zhang, “On matrix exponential distributions,” *Adv. Appl. Probab.*, vol. 39, no. 1, pp. 271–292, Mar. 2007.

[28] P. Buchholz and M. Telek, “Stochastic Petri nets with matrix exponentially distributed firing times,” *Perform. Eval.*, vol. 67, no. 12, pp. 1373–1385, Dec. 2010.

[29] N. Akar and K. Sohraby, “Infinite- and finite-buffer Markov fluid queues: A unified analysis,” *J. Appl. Probab.*, vol. 41, no. 2, pp. 557–569, Jun. 2004.

[30] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[31] H. E. Kankaya and N. Akar, “Solving multi-regime feedback fluid queues,” *Stochastic Models*, vol. 24, no. 3, pp. 425–450, Aug. 2008.

[32] N. Akar and K. Sohraby, “System-theoretical algorithmic solution to waiting times in semi-Markov queues,” *Perform. Eval.*, vol. 66, no. 11, pp. 587–606, Nov. 2009.

[33] H. C. Tijms, *A First Course in Stochastic Models*. West Sussex, U.K.: Wiley, 2003.

Nail Akar (Member, IEEE) received the B.S. degree from Middle East Technical University, Turkey, in 1987, and the M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1989 and 1994, respectively, all in electrical and electronics engineering. From 1994 to 1996, he was a Visiting Scholar and a Visiting Assistant Professor with the Computer Science Telecommunications Program, University of Missouri–Kansas City, USA. He joined the Technology Planning and Integration Group, Long Distance Division, Sprint, Overland Park, Kansas, in 1996, where he was a Senior Member of technical staff position, from 1999 to 2000. Since 2000, he has been with Bilkent University, where he is currently as a Professor with Electrical and Electronics Engineering Department. He visited the School of Computing, University of Missouri–Kansas City, as a Fulbright Scholar, in 2010, for a period of six months. His research interests include the performance analysis of computer and communication systems and networks, queuing models and tools, wireless networks, the Internet of Things.

Ozancan Doğan received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2018, where he is currently pursuing the M.S. degree. His current research interest includes the performance modeling of age of information.

Eray Unsal Atay is currently pursuing the bachelor’s degree in electrical and electronics engineering and mathematics with Bilkent University. His current research interests include probability theory and queuing systems. He received the Bronze Medal at the 57th International Mathematical Olympiad, Hong Kong.