EXACT DISTRIBUTION OF AGE OF INFORMATION (AOI) AND PEAK AOI IN SINGLE-SOURCE AND MULTI-SOURCE STATUS UPDATE SYSTEMS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Ozancan Doğan

July 2021

Exact Distribution of Age of Information (AoI) and Peak AoI in Single-Source and Multi-Source Status Update Systems

By Ozancan Doğan

July 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Nail Akar(Advisor)

Sinan Gezici

Mehmet Akif Yazıcı

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

EXACT DISTRIBUTION OF AGE OF INFORMATION (AOI) AND PEAK AOI IN SINGLE-SOURCE AND MULTI-SOURCE STATUS UPDATE SYSTEMS

Ozancan Doğan
M.S. in Electrical and Electronics Engineering
Advisor: Nail Akar
July 2021

In this thesis, we first study bufferless and single-buffer single-source queueing models of a status update system with various accompanying buffer management schemes. Next, we study the bufferless multi-source queueing model of a status-update system with probabilistic preemption. For both single-source and multi-source queueing models, we obtain the exact distributions of the associated Age of Information (AoI) and Peak Age of Information (PAoI) processes. For this purpose, we propose a Markov Fluid Queue (MFQ) model for both scenarios out of which the exact AoI distributions are derived. The numerical solution obtained from the proposed model provides the distributional expressions in matrix-exponential form out of which one can easily obtain their high order moments. We validate the accuracy of our proposed algorithm by comparing our results with simulations and also existing results in the literature in certain sub-cases. Numerical results are presented to provide engineering insight on how state update systems need to be configured and operated.

Keywords: Age of Information, Peak Age of Information, Status Update Systems, Markov Fluid Queues.

ÖZET

TEK KAYNAKLI VE ÇOK KAYNAKLI DURUM GÜNCELLEME SİSTEMLERİNDE BİLGİ YAŞI VE ZİRVE BİLGİ YAŞININ KESİN DAĞILIMI

Ozancan Doğan
Elektrik Elektronik Mühendisliği, Yüksek Lisans
Tez Danışmanı: Nail Akar
Temmuz 2021

Bu tezde, tek kuyruklu ve kuyruksuz tek kaynaklı durum güncellemesi sisteminin kuyruk modelini çeşitli kuyruk yönetimi yöntemleri altında çalıştık. Ayrıca, olasılıksal boşaltmalı ve kuyruksuz durum güncellemesi sisteminin çoklu kaynaklı kuyruk modelini de çalıştık. Tekli ve çoklu kaynaklı kuyruk modellerinin ikisi için de, AoI ve PAoI süreçlerinin kesin dağılımlarını elde ettik. Bu amaçla, Markov Akışkan Kuyrukları modelini her iki model için de önerdik ve kesin dağılımları bu model aracılığıyla elde ettik. Önerilen modelle elde edilen sayısal çözüm bize kesin dağılımları matris üstel formda elde etmemizi sağladı ve bu form sayesinde yüksek mertebedeki momentleri ve birikimli dağılımları kolayca hesaplayabildik. Önerdiğimiz yöntemlerin başarısını çeşitli alt durumlar için simülasyonlarla ve literatürde şu ana kadar bulunan sonuçlarla doğruladık. Sayısal sonuçları durum güncellemesi yapan sistemlerin konfigürasyonu ve operasyonu için mühendislik içgörüsü sağlamak için sunduk.

Anahtar sözcükler: Bilgi Yaşı, Zirve Bilgi Yaşı, Durum Güncelleme Sistemleri, Markov Akışkan Kuyrukları.

Acknowledgement

Firstly, I want to express my deepest gratitude to my advisor Prof. Nail Akar. He provided invaluable encouragement and support throughout my graduate studies. In addition, his enthusiasm and passion towards science has been invaluable source of inspiration.

I would also like to thank Information and Communication Technologies Authority (ICTA) and Vodafone for funding my study under 5G and Beyond scholarship program.

Finally, I would like to thank my family members for their unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Age of Information

Consider a real-time status update system that contains an information source, a server, and a destination (or a monitor). Status updates from the information source are sent and received by the server where the received updates are processed to be sent to the destination. A variety of applications require the delivery of real-time status updates to a destination from information sources. Wireless sensor networks for environmental monitoring that has been used for surveillance and monitoring of agriculture [5] or vehicular ad-hoc networks that provide communication between nearby vehicles [6] are examples of these applications. Timeliness of information could prevent collisions or shorten fire response time in these particular examples. Therefore, keeping the destination with the latest information is one of the main goals for these systems. To evaluate the timeliness of status updates at the destination, a new metric called Age of Information (AoI) is defined in [7],[8]. The AoI process is a random process defined as $\Delta(t) = t - u(t)$ where $u(t)$ is the generation time of the latest received update at the destination. In [9], a related performance metric peak Age of Information (PAoI) is introduced, which is defined as the value of AoI just before the reception of an information update.

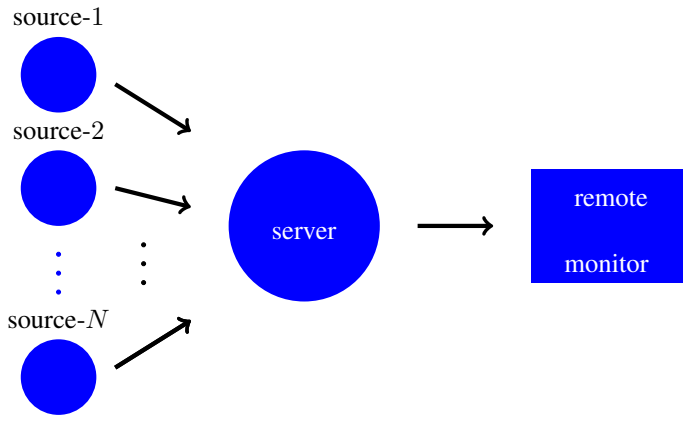

Figure 1.1: N information sources transmit status update packets to a remote monitor via a server.

For the AoI framework, queueing models are widely used to model communication systems in the literature. By adjusting network parameters such as scheduling policy of the server or the generation frequency of status updates, minimization and optimization of AoI is possible. Therefore, the analysis of AoI in queueing models to find its distribution and moments received considerable attention.

The single-server model in Fig. 1.1 has been extensively studied in the literature. There are many variations of this model that have been studied in the recent literature depending on the number of sources (single vs multiple), the existence of transmission errors and retransmission probabilities, the distributions assumed for inter-arrival and service times, the queue capacity of the system, the buffer management mechanism, the scheduling discipline, etc. In this thesis, we are particularly interested in two specific problems whose certain variations have also been studied in the literature. The first problem is the continuous-time single-source bufferless queue with probabilistic preemption and single-source single buffer queue with probabilistic replacement. As the second problem, we study the multi-source bufferless queue with probabilistic preemption, under transmission errors and retransmissions. We also studied a discrete-time queueing model in [10], however this study is not included to preserve the coherence in this thesis.

1.2 Contributions of the Thesis

The main contributions of the thesis can be expressed as following:

- We model single source and multi-source queues with Markov Fluid Queues (MFQs). Expressions for AoI and PAoI distributions in matrix exponential form are obtainable by solving the steady-state distribution of the proposed MFQs with existing stable and efficient numerical algorithms in the literature. We also show that distributions, moments, and tail probabilities of AoI and PAoI are easily obtainable from these expressions.
- We present a unifying framework for single-source systems allowing probabilistic preemption and replacement. A number of studied queueing models in the literature could be shown to be subcases of our general bufferless and single buffer models with certain preemption and replacement probabilities. We also show that probabilistic preemption with optimum preemption parameter could outperform existing models in certain scenarios.
- For multi-source systems, we introduce the so-called probabilistic preemption matrix which makes it possible to define distinct preemption probabilities among sources. It is also shown that existing global-preemption, no-preemption, and self-preemption policies are sub-cases of our unifying model. Furthermore, we define distinct error and transmission probabilities for each source. Again, we show that probabilistic preemption can be shown to be optimum in certain scenarios.

1.3 List of Publications

Single-source and multi-source queue models presented in Chapter 3 and 4, and numerical results in Chapter 5 are based on the following publications.

- 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, 2020

- O. Dogan and N. Akar, "The multi-source preemptive M/PH/1/1 queue with packet errors: Exact distribution of the age of information and its peak," *arXiv preprint arXiv:2007.11656*, 2020

1.4 Related Work

Analysis of AoI and PAoI in queueing models, and their optimization and minimization have been widely studied in the literature. Analysis of AoI in multi-hop networks has received considerable attention [13], [14], [15]. In the most of the literature and this thesis, single-hop networks are studied, as shown in Fig. 1.2.

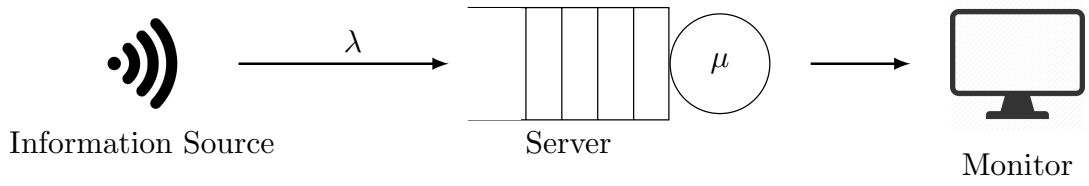

Figure 1.2: Single-hop queueing model

Single-hop networks are mainly categorized by the number of sources, number of servers, distributions of arrival and service processes, and scheduling disciplines. Even though some studies have been done for multi-server case [16],[17], single server was the main framework for most studies. In [7], first-come first-served (FCFS) M/M/1 queue is studied. In this study, mean AoI is obtained through integrating decomposed geometric parts. Preemptive and non-preemptive last-come first-served (LCFS) M/M/1 queues are studied by [18]. [19] obtains expressions for the average age for two different M/G/1/1 queues, one with preemption and the other one with blocking. Expressions for the mean AoI and mean PAoI for preemptive and non-preemptive LCFS queues with gamma distributed service

times are obtained in [20]. Deterministic service times are also approximated with Erlang distributions in this study. Preemptive scheduling is also referred as M/M/1/1* or preemption in service and incoming packet is allowed to preempt the packet in service. Non-preemptive scheduling is also referred as M/M/1/2* or preemption in waiting and incoming packet is allowed to preempt packet in the buffer. Multi-source FCFS, preemptive LCFS and non-preemptive LCFS M/M/1 queues are modelled as stochastic hybrid systems (SHS) and expressions for mean AoI are obtained by [3]. The average AoI is obtained for M/M/1/m blocking queue by their revisit to the topic [21] which is also an extension of their single-source work [19]. SHS technique is also applied to a tandem of preemptive LCFS queues [22] and a tandem of non-preemptive FCFS queues [23]. Derivation of exact expressions for the average AoI for multi-source M/M/1 queueing model, approximate expressions for multi source M/G/1 model are provided by [24]. Mean AoI and PAoI expressions for the M/M/1/1 queue and mean PAoI expression for M/M/1/2* queue are obtained by [9]. To improve the performance, packet deadlines for the M/M/1/2 queue are introduced by [25] and in this system, the packet in the queue is discarded if its wait time exceeds a deadline. The reference [2] obtains the Laplace transform expressions for AoI in FCFS, preemptive LCFS and non-preemptive LCFS M/GI/1 and GI/M/1 queues where GI denotes independent and identically distributed general distributions. Retransmission of errored packets is found to increase performance in LCFS systems in [26]. Furthermore, [17] finds that Last Generated First Served (LGFS) policy optimizes the age along with throughput and delay for infinite buffer queueing systems.

Discrete-time AoI models are also studied in te literature but to a lesser extent than continuous-time models. The discrete-time FIFO G/Ber/1 queue is studied in [27] where the authors study the average AoI and PAoI. In [28], the authors obtain the average AoI from Z-transforms for general FCFS and preemptive LCFS queue disciplines and they apply these results to Geo/Geo/1/1 queues. Explicit expressions for the FCFS Ber/G/1 and Ber/G/1 queues with vacations are derived by [29]. The LCFS G/G/∞ is studied as well in this study. The reference

[30] studies the optimization of AoI in wireless networks by combining queueing theory and stochastic geometry. The exact distributions of AoI and PAoI are obtained in matrix-geometric form for Non-Preemptive Bufferless (NPB) ii) Preemptive Bufferless (PB) iii) Non-Preemptive Single Buffer with Replacement (NPSBR) queue disciplines where Bernoulli arrivals and general discrete phase-type service times are assumed [10].

1.5 Thesis Organization

The rest of the thesis is organized as follows. First, we present the preliminaries on phase-type distributions and Markov fluid queues in Chapter 2. We study the AoI and PAoI processes in single-source and multi-source queues with proposed models in Chapter 3 and 4, respectively. In Chapter 5, we provide numerical examples to validate our proposed models' accuracy and present results obtained with our proposed models. Finally, we conclude in Chapter 6.

Chapter 2

Preliminaries

Upper case and lower case bold letters denote the real valued matrices and vectors, respectively. The elements of matrix A is denoted by $A_{i,j}$ for $i, j \in 1, 2, ..., n$. The vectors $\mathbf{0}_{k\times\ell}$, $\boldsymbol{I}_m$, and $\mathbf{1}_n$ denote matrix of zeros with size $k \times \ell$, identity matrix with size m and column vector of ones with size n, respectively. For a vector $\boldsymbol{\alpha}$, α_j denotes the jth element. Kronecker product and diagonal concatenation of two matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{p\times q}$ is denoted by $\boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{R}^{mp\times nq}$ and $\mathbf{diag}\{\boldsymbol{A}, \boldsymbol{B}\}$, respectively. The concatenation of two row vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is denoted by the notation $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$. The unit step and Dirac delta functions are denoted by $u(x)$ and $\delta(x)$, respectively.

2.1 Phase-type Distributions

Consider a Markov process that has $m + 1$ states where the first m states are transient states and the last state is the absorbing state. And the initial probability vector α describes the initial probability of starting in any of states defined in state space S. The distribution of the time until the absorption occurs is called Phase-type (PH-type) distribution. Various probability distributions can be fit to phase-type distribution via expectation-maximization algorithm [31]. A PH-type

random variable can be represented by the pair (S, α). The generator matrix for Markov Chain Q is given as:

$$Q = \left[\begin{array}{c:c} S & \boldsymbol{\nu} \\ \hdashline \mathbf{0} & 0 \end{array}\right], \tag{2.1}$$

S is size of $m \times m$ and ν is size of $m \times 1$. S_{ij} represents the transition rate from state i to j where both states are transient and v_i represents the transition rate from i to absorbing state. Since row sums of generator matrix should be equal to 0, $\boldsymbol{\nu} = -S1$. The cdf and pdf of a PH-type random variable are given as:

$$\begin{aligned} F_X(x) &= (1 - \boldsymbol{\sigma} e^{Sx}\mathbf{1})u(x), \\ f_X(x) &= -\boldsymbol{\sigma} e^{Sx} S\mathbf{1}u(x) + \sigma_0\delta(x). \end{aligned} \tag{2.2}$$

The more general form of PH-type distributions are called matrix exponential distributions [32]. Matrix exponential distributed random variable is also described by the pair (σ, S). Different from PH-type distributions, there is no restrictions on σ and S. The moments of PH-type and ME-type random variables are given as:

$$E[X^i] = i!\boldsymbol{\sigma}(-S)^{-i}\mathbf{1},\ i = 1, 2, \ldots. \tag{2.3}$$

Probability density functions of Erlang and hyperexponential distributions are given in order as:

$$f(x; m, \lambda) = \frac{\lambda^m x^{m-1} e^{-\lambda x}}{(m-1)!} \tag{2.4}$$

$$f(x) = \sum_{i=1}^{m} p_i \lambda_i e^{-\lambda_i x} \tag{2.5}$$

These two distributions can be easily expressed with PH-type distributions. Fig. 2.1 and 2.2 show the sample diagram for the underlying Markov chain of an Erlang distribution and hyperexponential distribution, respectively. Generator matrix and initial probability vector for the Markov chain of Erlang distribution (Q_1, α_1) and hyperexponential distribution (Q_2, α_2) are given as:

$$Q_1 = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & \cdots & 0 & 0 & 0 \\ & \cdots & & & & \cdots & \\ 0 & 0 & 0 & \cdots & 0 & -\lambda & \lambda \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\lambda \end{bmatrix}, \alpha_1 = (1, 0, 0..., 0) \tag{2.6}$$

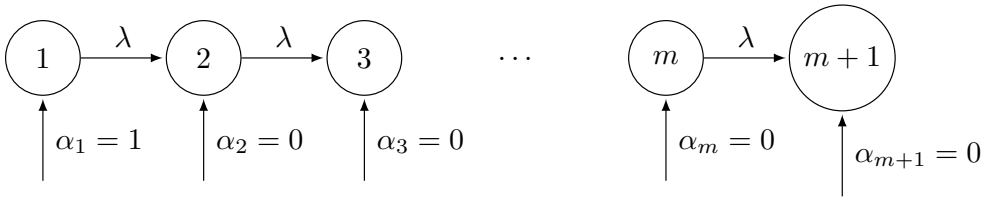

Figure 2.1: Underlying Markov chain for Erlang distribution

$$Q_2 = \begin{bmatrix} -\lambda_1 & 0 & 0 & \cdots & 0 & 0 & \lambda_1 \\ 0 & -\lambda_2 & 0 & \cdots & 0 & 0 & \lambda_2 \\ & \cdots & & & & \cdots & \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_m & \lambda_m \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\lambda \end{bmatrix}, \alpha_2 = (p_1, p_2 ... p_m, 0) \tag{2.7}$$

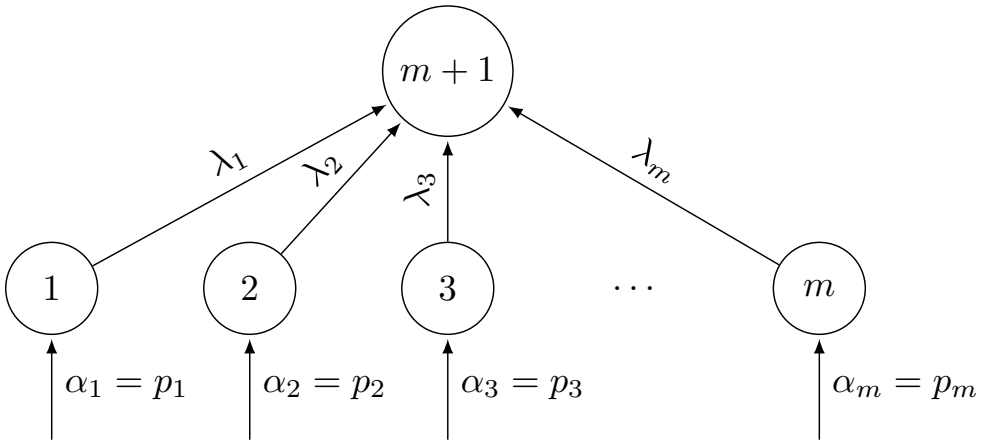

Figure 2.2: Underlying Markov chain for hyperexponential distribution

ME distribution in the form of (2.8) can be written in the form of (2.2) with given method in [33]. This type of substitution is required in the analysis of single-buffer single-source queue where the result of General Markov Fluid Queue (GMFQ) is required to be placed on the generator matrix of another GMFQ.

$$f_X(x) = \boldsymbol{g} e^{Ax} \boldsymbol{h} u(x) + \sigma_0 \delta(x), \tag{2.8}$$

Lemma 1. *Let X be a rv with a pdf $f_X(x)$ which has following matrix exponential form:*

$$f_X(x) = \boldsymbol{g}e^{Ax}\boldsymbol{h}u(x) + \sigma_0\delta(x),$$
$$E[X^i] = (-1)^{i+1}i!\boldsymbol{g}A^{-(i+1)}\boldsymbol{h}, \tag{2.9}$$

for $i = 1, 2, \ldots,$ with A having the size n and $\int_{-\infty}^{\infty} f_X(x)\,dx = -\boldsymbol{g}A^{-1}\boldsymbol{h} + \sigma_0 = 1$. We can find a vector $\boldsymbol{\sigma}$ and a matrix S such that $X \sim ME(\boldsymbol{\sigma}, S)$ with the same order n.

Proof. When $h \neq 0$, a nonsingular matrix M exists that satisfies $M\mathbf{1} = \boldsymbol{v}$ where $\boldsymbol{v} = A^{-1}\boldsymbol{h}$. Firstly, we notice that A's eigenvalues are in the open left-half plane since X has a legitimate pdf. Moreover, $\boldsymbol{v}$ should have at least one non-zero element since A is nonsingular and h is not vector of zeros.

To construct matrix M satisfying $M\mathbf{1} = \boldsymbol{v}$ and non-singularity, we apply the following rule. Let k be the least index $v_k \neq 0$, we set $M_{k,k} = v_k$ and other elements on the k'th row to 0. If $i \neq k$ and $q_i \neq 0$, we set $M_{i,i} = v_i$ and all other elements in the selected row to 0. If $i \neq k$ and $q_i \neq 0$, we set $M_{i,i} = 1$ and all other elements in the selected row to -1. Certainly, the constructed matrix is nonsingular and satisfies $M\mathbf{1} = \boldsymbol{v}$. It can be easily shown that $f_X(x)$ can be expressed in form of 2 by choosing $\boldsymbol{\sigma} = \boldsymbol{g}M$, $S = M^{-1}AM$ so that $X \sim ME(\boldsymbol{\sigma}, S)$.

□

2.2 Markov Fluid Queues

Fluid queues are special kind of queues where the rate of change in the fluid level in the buffer is varied with respect to the state of the Markov Chain. A Markov Fluid Queue is a fluid queue which is described by a joint Markovian process $X(t) = (X_f(t), X_m(t))$ where $X_f(t)$ describes the fluid level in the buffer $(0 \leq X_f(t) \leq \infty)$ and $X_m(t)$ (underlying process) describes the net rate of the

change in the fluid level. $X_m(t)$ is continuous time markov chain (CTMC) where $X_m(t)$ which has its own state-space and generator matrix Q. Net rate of change in the fluid level is determined by the phase of the underlying process $X_m(t)$. When $X_m(t)$ is in phase i, change is described by r_i which is also called drift. Drift matrix R is a diagonal matrix of drifts consists of drifts in all phases such that $\boldsymbol{R} = \mathbf{diag}\{r_1, r_2, \ldots, r_n\}$. We are interested with a special kind of MFQ where buffer capacity is infinite ($0 \leq X_f(t) \leq \infty$). In our case behavior of the underlying process depends on the fluid level in the buffer. When $X_f(t) > 0$, $X_m(t)$ behaves according to the infinitesimal generator Q, and when $X_f(t) = 0$ the generator is denoted by $\tilde{Q}$. Drift matrix for both cases equals to R. This type of MFQ is called GMFQ and is characterized by the triple $(Q, \tilde{Q}, R)$. Our interest is to find steady-state joint density pdf vector (2.10) and probability mass accumulation vector (pma) at zero (2.11).

$$\boldsymbol{f}(x) = [f_1(x), f_2(x), \ldots, f_n(x)]\,, \tag{2.10}$$

where $f_i(x) = \lim_{t\to\infty} \frac{d}{dx} \Pr\{X_f(t) \leq x, X_m(t) = i\}$ for $x > 0$.

$$\boldsymbol{c} = [c_1, c_2, \ldots, c_n]\,, \tag{2.11}$$

where $c_i = \lim_{t\to\infty} \Pr\{X_f(t) = 0, X_m(t) = i\}$.

Based on the numerical algorithm in [34], the Alg. 1 that explains the pseudo-code to obtain joint density (2.10) and mass accumulation (2.11) vectors has been provided by [11].

Apart from joint density and mass accumulation vectors, we also define a quantity expresses the fluid level's steady-state conditional density before transitioning from a state that is a member of a particular subset $\mathcal{S}_0 \subset \mathcal{S}$ to a state j which is not a member of this subset. Closed-form expression of conditional steady state density $g_j^{\mathcal{S}_0}(x)$ is given in terms of steady-state joint pdf vector of MFQ $X(t)$ is provided by the following theorem.

Theorem 1. *Let $\boldsymbol{X}(t) \sim GMFQ(Q, \tilde{Q}, R)$ with order n and its steady-state joint pdf $\boldsymbol{f}(x)$ is given as in (2.13). We choose a set of states $\mathcal{S}_0 \subset \mathcal{S}$ such that the selected states in $\mathcal{S}_0$ dont have probability mass at zero. Then, the steady-state*

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

1: **function** STEADY-STATE$(Q, \tilde{Q}, R, n, a, b)$ ▷ n is the size of these matrices

2: Step 1: Find an orthogonal matrix P that puts the matrix QR^{-1} into the following form:

$$P^T QR^{-1}P = \left[\begin{array}{c:c} F_{a\times a} & * \\ \hdashline \mathbf{0} & A_{b\times b} \end{array}\right], P^T = \left[\begin{array}{c:c} * \\ \hdashline H_{b\times n} \end{array}\right] \tag{2.12}$$

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 (2.12); see [35],[36] and the references therein for the Schur form and its numerical stability.

3: 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 $\tilde{Q}$.

4: Step 3: Write $\boldsymbol{c} = [\mathbf{0}, \boldsymbol{d}]$ and the steady-state joint pdf vector as

$$\boldsymbol{f}(x) = \boldsymbol{g}e^{Ax}Hu(x) + \boldsymbol{c}\delta(x), f_i(x) = \boldsymbol{g}e^{Ax}\boldsymbol{h}_i u(x) + c_i\delta(x), \tag{2.13}$$

where $\boldsymbol{h}_i$ denotes the i^{th} column of H.

5: **end function**

conditional density $g_j^{\mathcal{S}_0}(x)$ is given in terms of the steady-state density vector $\boldsymbol{f}(x)$ by

$$g_j^{\mathcal{S}_0}(x) = \frac{\sum\limits_{i\in\mathcal{S}_0} f_i(x)Q_{i,j}}{\int\limits_0^\infty \sum\limits_{i\in\mathcal{S}_0} f_i(x')Q_{i,j}dx'} = \boldsymbol{g}'e^{Ax}\boldsymbol{h}\ u(x), \tag{2.14}$$

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 firstly write cumulative distribution as:

$$\begin{aligned} &\lim_{\substack{t\to\infty \\ \Delta t\to 0}} \Pr\{X_f(t) \le x \mid X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\} = \\ &\lim_{\Delta t\to 0} \frac{\lim\limits_{t\to\infty} \Pr\{X_f(t) \le x, X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\}}{\lim\limits_{t\to\infty} \Pr\{X_m(t) \in \mathcal{S}_0, X_m(t+\Delta t) = j\}}. \end{aligned} \tag{2.15}$$

Since Δt approaches 0, we can write the denominator of (2.15) as:

$$= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\} \cdot \gamma_i(t),$$

$$= \sum_{i\in\mathcal{S}_0} Q_{i,j}\Delta t \cdot \int_0^\infty f_i(x')dx' = \Delta t \int_0^\infty \sum_{i\in\mathcal{S}_0} f_i(x')Q_{i,j}dx'.$$

where $\gamma_i(t) = \Pr\{X_m(t) = i\}$. At the same time, as Δt approaches 0, the numerator of (2.15) can be written as:

$$= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x, X_m(t+\Delta t) = j \mid X_m(t) = i\} \cdot \gamma_i(t),$$

$$= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x \mid X_m(t) = i\} \cdot \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\} \cdot \gamma_i(t),$$

$$= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \frac{\Pr\{X_f(t) \le x,\ X_m(t) = i\}}{\Pr\{X_m(t) = i\}} \cdot \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\} \cdot \gamma_i(t),$$

$$= \lim_{t\to\infty} \sum_{i\in\mathcal{S}_0} \Pr\{X_f(t) \le x, X_m(t) = i\} \cdot \Pr\{X_m(t+\Delta t) = j \mid X_m(t) = i\},$$

$$= \Delta t \sum_{i\in\mathcal{S}_0} Q_{i,j} \int_0^x f_i(x')dx'.$$

Derivative of the expression (2.15) gives the left-hand side of (2.14). Numerator of the first expression equals to $\boldsymbol{g}e^{A\boldsymbol{x}}\boldsymbol{h}\ u(x)$ and the denominator equals to normalization constant $-\boldsymbol{g}A^{-1}\boldsymbol{h}$ so that expression has a legitimate pdf. With the choice of $\boldsymbol{g'} = \boldsymbol{g}/(-\boldsymbol{g}A^{-1}\boldsymbol{h})$, we can finally express $g_j^{\mathcal{S}_0}(x)$ as the right-hand side of (2.14).

□

Chapter 3

Single-Source System

3.1 The General Model

To model the single-source update system, we consider a system model containing a single source, a server, and a monitor as shown in Fig. 3.1.

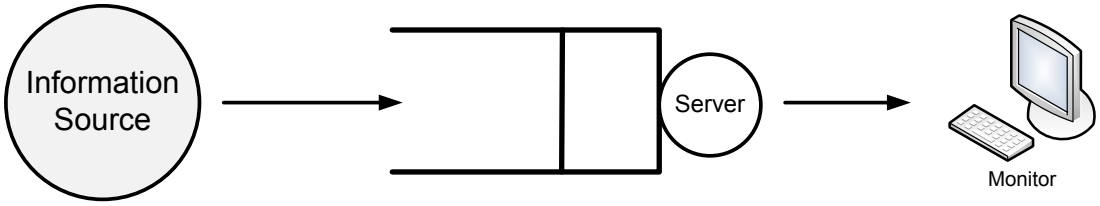

Figure 3.1: An information source sending status update messages through a queue to a remote monitor

The information source send its status updates to a server where the processing of these update takes place. When the service time of an update is over, it is displayed on the monitor. Successful and unsuccessful packets with arriving order j are denoted by the notation t_j and t_j^*, respectively. The interarrival time between packets are denoted by Λ_j. Interarrival times are assumed to be independent and identically distributed with probability density function(pdf)

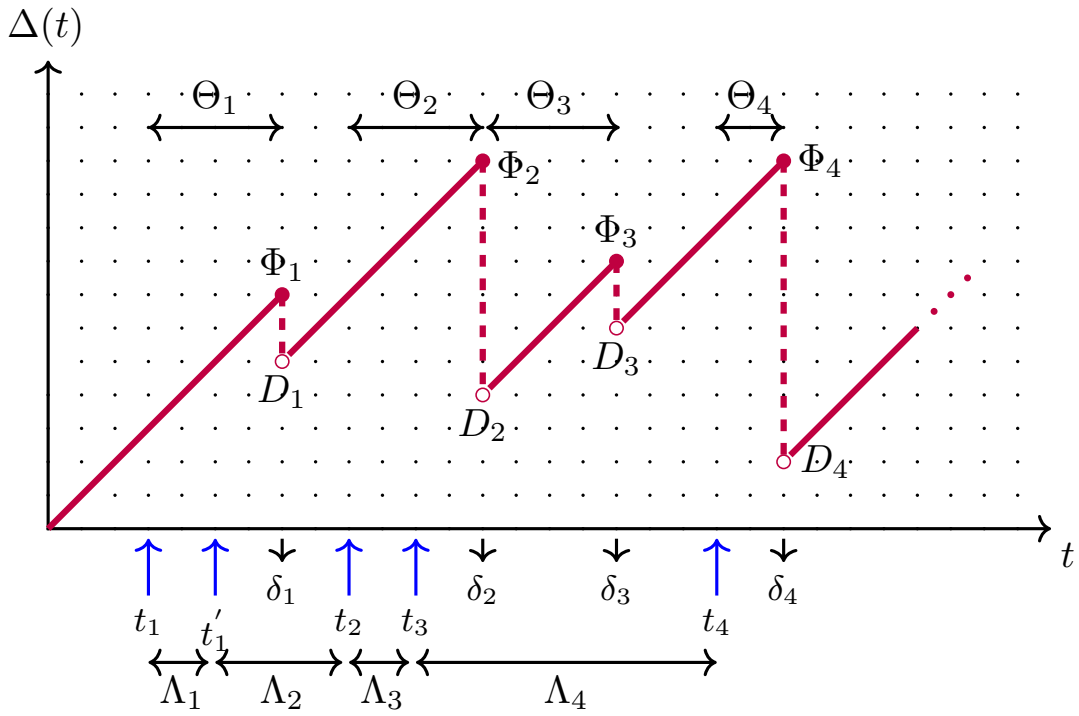

Figure 3.2: An illustration of the sample path of AoI and PAoI processes for single source and single buffer system.

$f_\Lambda(\cdot)$, cumulative distribution function(cdf) $F_\Lambda(\cdot)$. Mean and variance of Λ are $1/\lambda$ and σ_Λ^2, respectively and squared coefficient of variation (scov) is defined as $c_\Lambda^2 = \sigma_\Lambda^2\lambda^2$. Service time of a packet j is denoted by Θ_j and it is i.i.d. with pdf $f_\Theta(\cdot)$, cdf $F_\Theta(\cdot)$ mean $1/\mu$ and scov $c_\Theta^2 = \sigma_\Theta^2\mu^2$. System time is the sum of queue wait time W_j and service time of a jth successful packet and denoted by D_j. Let δ_j denotes the reception time of jth packet then, $D_j = \delta_j - t_j$. For multi-source case, the same notation applies with an additional superscript which denotes the source of the packet.

Fig. 3.2 shows a sample path of AoI and PAoI processes denoted by $\Delta(t)$ and Φ_j, respectively in a single-buffer and single-source system. In this example, $\Delta(0) = 0$, and it increases linearly until service of packet 1 is over at $t = 6$, then $\Phi_1 = 7$ and $D_1 = 4$ which is equal to the service time of the first successful packet. The second arriving packet is preempted by the second successful packet at $t = 9$ and the third successful packet is taken to buffer at $t = 9$. When the service of the second successful packet is over at $t = 12$, $\Phi_2 = 10$ and $D_2 = 4$ which is the

service time of the the second successful packet. When the service time of the third packet is over at $t = 16$, $\Phi_3 = 9$ and $D_3 = 7$. In this case, system time (D_3) is the sum of queue wait time and service time of the third packet. And finally, at $t = 21$ service of the fourth successful packet is over, then $\Phi_4 = 10$ and $D_2 = 2$ which is the service time of fourth successful packet.

For each cycle, the process $\Delta(t)$ increases linearly until a packet reception occurs where the process equals to peak value Φ_j and it is simply a sample of PAoI process and it drops to D_j which is the system time of last successful packet for both examples. Our main motivation is to find cdfs of AoI and PAoI processes given by Eq. 3.1 for single-source system with employing the theory of Markov Fluid Queues (MFQ) [37], [38], [39].

$$F_\Delta(x) = \lim_{t\to\infty} \Pr\{\Delta(t) \leq x\}, F_\Phi(x) = \lim_{j\to\infty} \Pr\{\Phi_j \leq x\}, \tag{3.1}$$

3.2 Bufferless Queues

In this subsection, we are interested in finding the exact distributions of AoI and PAoI processes in the $PH/PH/1/1/P(p)$ queue, where the maximum number of packets allowed in the system is 1. In this queueing model, the information packet in service is preempted with probability p by a new arrival. When $p = 0$, queueing system turns to be FCFS $PH/PH/1/1$ queue, and when $p = 1$, queueing system turns to be preemptive LCFS queue. For this queueing model, the interarrival 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 $\boldsymbol{\nu} = -S\mathbf{1}, \sigma_0 = 1 - \boldsymbol{\sigma}\mathbf{1} = 0$. To be able to obtain the exact distributions of AoI and PAoI, we need to construct a GMFQ process $X(t)$ which consists of infinitely many cycles and these cycles begin with the arrival of the first successful packet into the system and ends with the successful reception of the second packet.

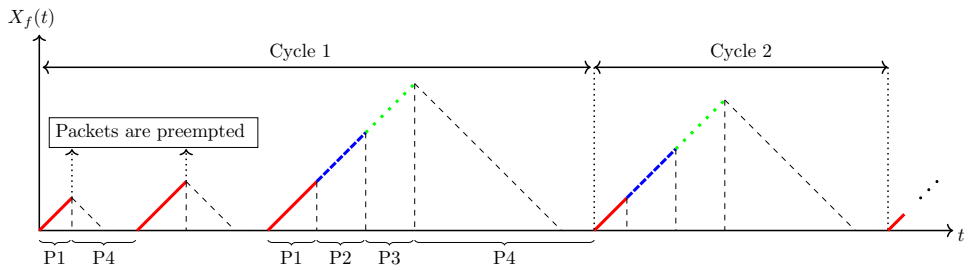

Figure 3.3: Sample path of the fluid level process $X_f(t)$ where Pj denotes phase j, $j = 1, \ldots, 4$.

Fig. 3.3 describes the sample path for the fluid level process of $X_f(t)$. Each cycle consists of four phases. The first phase (solid red curve) is used to construct the system time which is the service time of the last successful packet. Since we aim to obtain system time of successful packets, unsuccessful packets are ignored by the system's design. The second phase (dashed blue curve) is used to construct the interarrival time after successful reception of the previous packet. The third phase (dotted green curve) represents the time of service process of the current packet which could be interrupted many times. In the sample path given in Figure 3.3, first two packets are preempted so they will not be taken into consideration since they are not successful. If a packet is interrupted before its service is completed, a transition from phase 1 to 4 (dashed black curve) occurs. After a successful reception, transition from phase 1 to 2 occurs. In phase 2, the system is waiting for the arrival of a new packet. This phase lasts until the arrival of a new packet. Phase 3 takes place until service time of the packet is over where many preemptions could take place. If any preemption occurs, the arrival and service processes are restarted. When the service is over, the system transitions into phase 4 where the fluid level is brought to level 0 with a drift rate -1. When fluid level becomes 0, it stays at the 0 fluid level with arbitrary time since masses at 0 are not included in the calculation.

Each phase consists of several states that we call this set of states as $\mathcal{S}_n, n = 1, ..., 4$. The state-space $\mathcal{S}$ of the modulating process $X_m(t)$ can be written as

$\mathcal{S} = \bigcup_{n=1}^{4} \mathcal{S}_n$. Where $\mathcal{S}_n$'s are given as the following:

$$\begin{aligned} \mathcal{S}_m &= \{(i^{(m)}, j^{(m)})\},\ i^{(m)} = 1, \ldots, k,\ j^{(m)} = 1, \ldots, \ell, \\ \mathcal{S}_2 &= \{i^{(2)}\},\ i^{(2)} = 1, \ldots, k, \end{aligned} \tag{3.2}$$

for $m = 1, 3$ and $\mathcal{S}_4 = \{0\}$. For phases 1 and 3, $i^{(m)}$ and $j^{(m)}$ keep track of the state of interarrival and service times, respectively. Characterizing matrices $(Q, \tilde{Q}, R)$ of the GMFQ process $X(t)$ are presented as;

$$Q = \left[\begin{array}{c|c|c|c} Q_{11} & \boldsymbol{I}_k \otimes \boldsymbol{\nu} & \mathbf{0} & p\boldsymbol{\kappa} \otimes \mathbf{1}_\ell \\ \hline \mathbf{0} & T & \boldsymbol{\kappa} \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma}) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & Q_{33} & \mathbf{1}_k \otimes \boldsymbol{\nu} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{array}\right], \tilde{Q} = \left[\begin{array}{c|c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \boldsymbol{\tau} \otimes \boldsymbol{\sigma} & \mathbf{0} & \mathbf{0} & -1 \end{array}\right],$$

$$R = \mathbf{diag}\{\boldsymbol{I}_{2k\ell+k}, -1\}, \tag{3.3}$$

where

$$\begin{aligned} Q_{11} &= \boldsymbol{I}_k \otimes S + T \otimes \boldsymbol{I}_\ell + (1-p)(\boldsymbol{\kappa} \otimes \boldsymbol{\tau}) \otimes \boldsymbol{I}_\ell, \\ Q_{33} &= Q_{11} + p\boldsymbol{\kappa} \otimes (\mathbf{1}_\ell \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma})). \end{aligned} \tag{3.4}$$

Detailed description of each term in matrices $(Q, \tilde{Q})$ is given in Table 3.1.

Table 3.1: A detailed explanation of each term in Q and $\tilde{Q}$ for $PH/PH/1/1$ queue.

Term	Description
$T \otimes \boldsymbol{I}_\ell$	State of the interarrival time changes from i_1 to i_2 and system transitions from state $(i_1^{(m)}, j_1^{(m)})$ to state $(i_2^{(m)}, j_1^{(m)})$ with transition rate $T_{(i_1^{(m)}, i_2^{(m)})}$ for $m = 1, 3$.
$\boldsymbol{I}_k \otimes S$	State of the service time changes from $j_1^{(m)}$ to $j_2^{(m)}$ and system transitions from state $(i_1^{(m)}, j_1^{(m)})$ to state $(i_1^{(m)}, j_2^{(m)})$ with transition rate $S_{(j_1^{(m)}, j_2^{(m)})}$ for $m = 1, 3$.
$(1-p)(\boldsymbol{\kappa} \otimes \boldsymbol{\tau}) \otimes \boldsymbol{I}_\ell$	The new arrival does not preempt the packet in the service with probability $1-p$ and arrival time is restarted.
$\boldsymbol{I}_k \otimes \boldsymbol{\nu}$	The service is completed and the system transitions from state $(i_1^{(1)}, j^{(1)})$ to state $i_1^{(2)}$ with transition rate $\boldsymbol{\nu}_{j^{(1)}}$.
$p\boldsymbol{\kappa} \otimes \mathbf{1}_\ell$	The packet in the service is preempted and the system transitions from state $(i^{(1)}, j^{(1)})$ to state 0 with transition rate $\boldsymbol{\kappa}_{i^{(1)}}$.
T	State of the interarrival time changes and the system transitions from state $i_1^{(2)}$ to $i_2^{(2)}$ with transition rate $T_{(i_1^{(2)}, i_2^{(2)})}$.
$\boldsymbol{\kappa} \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma})$	A new arrival occurs, service and arrival process is restarted and we transition from state $i_1^{(2)}$ to state $(i_2^{(3)}, j^{(3)})$ with transition rate $\boldsymbol{\kappa}_{i_1^{(2)}} \boldsymbol{\tau}_{i_2^{(3)}} \boldsymbol{\sigma}_{j^{(3)}}$.
$p\boldsymbol{\kappa} \otimes (\mathbf{1}_\ell \otimes (\boldsymbol{\tau} \otimes \boldsymbol{\sigma}))$	The packet in the service is preempted by the new arrival and service and arrival process is restarted.
$\mathbf{1}_k \otimes \boldsymbol{\nu}$	The service process is completed and the system transitions from state $(i^{(3)}, j^{(3)})$ to state 0 with transition rate $\boldsymbol{\nu}_{j^{(3)}}$.
$\boldsymbol{\tau} \otimes \boldsymbol{\sigma}$	Service and arrival time restart and the system transitions from state 0 to state $(i^{(1)}, j^{(1)})$ with probability $\boldsymbol{\tau}_{i^{(1)}} \boldsymbol{\sigma}_{j^{(1)}}$.

The fluid level $X_f(t)$ at the beginning of phase 2 represents the D_j and the

fluid level just at the beginning of phase 4 represents the Φ_{j+1}. Therefore, the fluid level in the concatenation of phases 2 and 3 represents the sample cycle of AoI process and the value of the fluid level $X_f(t)$ just at the beginning of phase 4 represents a sample value of PAoI process. Since we can obtain the steady-state solution of $X(t)$, we can also obtain the exact distributions of the AoI and PAoI with the following theorem.

Theorem 2. *Consider the $GMFQ(Q, \tilde{Q}, R)$ with the characterizing matrices as defined in* (3.3) *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 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:*

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

$$f_\Phi(x) = \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{\ell} f_{(i^{(3)},j^{(3)})}(x)\nu_j}{\int\limits_0^\infty \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{\ell} f_{(i^{(3)},j^{(3)})}(x')\nu_j\,dx'}. \tag{3.5}$$

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. 3.3, 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 (3.5) 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. □

3.3 Single Buffer Queues

In this subsection, we are interested in finding the exact distributions of AoI and PAoI processes in the $M/PH/1/2/R(r)$ queue where the maximum number of packets allowed in the system is two, one in the buffer and one in the service. In this model, the information packet joins the queue if the buffer is empty otherwise this new arrival replaces the packet in the buffer with probability r. When $r = 0$ queueing system turns out to be FCFS $M/PH/1/2$ queue and when $r = 1$ the system turns out to be non-preemptive LCFS queue. In this queueing model, interarrival time is exponentially distributed with parameter λ and the service time $\Theta \sim PH(\boldsymbol{\sigma}, S)$ with order ℓ and $\boldsymbol{\nu} = -S\mathbf{1}, \sigma_0 = 1 - \boldsymbol{\sigma}\mathbf{1} = 0$. Since the system has buffer, system time(D_j) consists of queue wait and service time of the last successful packet. To be able to obtain pdf of queue wait time W, we need to construct a GMFQ process $Y(t)$ which is related to residual service time.

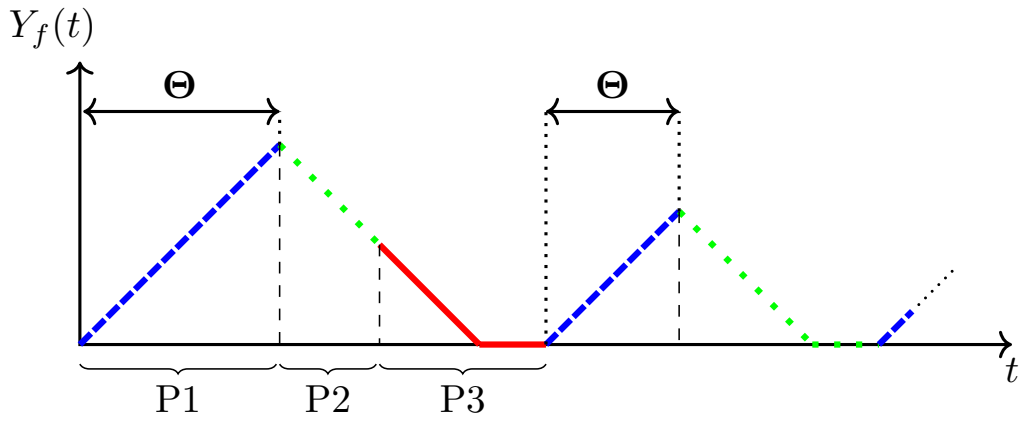

Figure 3.4: Sample path of the fluid level process $Y_f(t)$ corresponding to the residual service time. Pj denotes phase j, $j = 1, \ldots, 3$.

Fig. 3.4 describes the sample path for fluid level process $Y_f(t)$. The first phase (shown by the dashed blue curve) is used to construct the service time of the last successful packet. When the fluid level reaches equal to the service time, system transitions from phase 1 to phase 2 (shown by the dotted green curve) where the drift rate is minus one, in other words the fluid level is decreasing with a unit rate.

If any arrival occurs during phase 2, transition from phase 2 to phase 3 (shown by the solid red curve) occurs. When the fluid level becomes zero, the system is either in phase 2 or phase 3. If the system is in phase 2 (no arrival occurred), a new arrival is waited to make a transition to phase 1. If the system is in phase 3, the system waits exponentially distributed time with unit parameter to make a transition to phase 1.

Each phase consists of several states that we call this set of states as $\mathcal{S}_n, n = 1, 2, 3$. The state-space $\mathcal{S}$ of the modulating process $Y_m(t)$ can be written as $\mathcal{S} = \bigcup_{n=1}^{3} \mathcal{S}_n$. Where $\mathcal{S}_n$'s are given as the following:

$$\mathcal{S}_1 = \{1^{(1)}, \ldots, \ell^{(1)}\}, \mathcal{S}_2 = \{0\} \text{ and } \mathcal{S}_3 = \{1\}. \tag{3.6}$$

Characterizing matrices $(Q, \tilde{Q}, R)$ of the GMFQ process $Z(t)$ is presented as;

$$Q = \left[\begin{array}{c|c|c} S & \boldsymbol{\nu} & \mathbf{0} \\ \hline \mathbf{0} & -\lambda & \lambda \\ \hline \mathbf{0} & 0 & 0 \end{array}\right], \; \tilde{Q} = \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \lambda\boldsymbol{\sigma} & -\lambda & 0 \\ \hline \boldsymbol{\sigma} & 0 & -1 \end{array}\right], \tag{3.7}$$

$$R = \mathbf{diag}\{\boldsymbol{I}_l, -\boldsymbol{I}_2\} \tag{3.8}$$

Since the buffer is not empty in phase 3, mass at zero for the state in phase 3 does not represent the residual service time. Then, the fluid level in phase 2 when $Y_f(t) \geq 0$ and in phase 3 when $Y_f(t) > 0$ represents the residual service time. Steady-state joint pdf for states 0 and 1 can be written as,

$$f_i(x) = \boldsymbol{g} e^{Ax} \boldsymbol{h}_i u(x) + c_i \delta(x), \; i = 0, 1, \tag{3.9}$$

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. Then the pdf of queue wait time $f_W(x)$ can be obtained with the following theorem.

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* (3.9) *as follows:*

$$f_W(x) = \eta_1 \left(\boldsymbol{g} e^{(A - r\lambda \boldsymbol{I}_\ell)x} (\boldsymbol{h}_0 + r\boldsymbol{h}_1) u(x) + c_0 \delta(x) \right), \tag{3.10}$$

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 $\tilde{f}_i(x)$:

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

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

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

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 (3.10) completing the proof. □

Remark. *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.*

As we have done in bufferless queues, we need to construct GMFQ process $Z(t)$ which consists of infinitely many cycles and again they begin with the arrival of the first successful packet into the system and ends with the successful reception of the second packet.

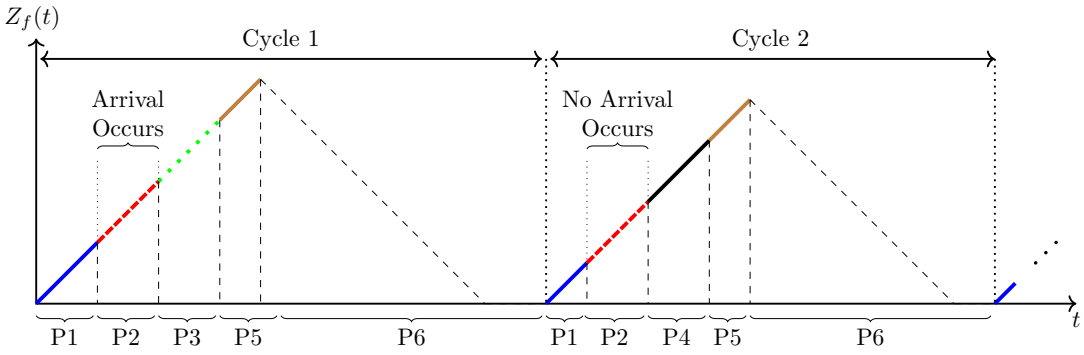

Figure 3.5: Sample path of the fluid level process $Z_f(t)$ where Pj denotes phase j, $j = 1, \ldots, 6$.

Fig. 3.5 describes the sample path for the fluid level process of $Z_f(t)$. Each cycle consists of six phases. The first phase (solid blue curve) is used to construct the queue wait time of the last successful packet. The second (shown by dashed red curve) and third (solid black curve) phase represents the service time of the last successful packet, where the first one represents the case the buffer is empty and the latter one represents the buffer is full. The combination of the first three phases is used to construct the system time (D_i) of the last successful packet. Phase 4 (dotted green curve) represents the time interval where the system is waiting for the arrival of the new packet before transition to phase 5 (solid brown curve). Phase 5 represents the service time of the second packet. From the beginning of phase 1 to the beginning of phase 6, the fluid level $Z_f(t)$ increases with drift rate 1. In the sample path given in Fig. 3.5 there are two example cycles. In both cycles, the system is waiting for the completion of the queue wait time. When the queue wait time is over, the system transitions from phase 1 to 2. In the first cycle one arrival occurs during the service of the first packet therefore the system transitions from phase 2 to phase 3. When the service is completed in phase 3, the system transitions from phase 3 to 5 since the packet in the buffer is taken to the service directly. In the second cycle, no arrival occurs during phase 2. When the service of the first packet is completed in phase 2, the system transitions from phase 2 to 4. In phase 4, the system is waiting for the arrival of a new packet. When the arrival occurs, the system transitions from

phase 4 to phase 5. When the service of the second packet is over, the system transitions to phase 6 where the fluid level $Z_f(t)$ is brought to level 0 with drift rate -1.

Each phase consists of several states that we call this set of states as $\mathcal{S}_n, n = 1, ..., 6$. The state-space $\mathcal{S}$ of the modulating process $Z_m(t)$ can be written as $\mathcal{S} = \bigcup_{n=1}^{6} \mathcal{S}_n$. Where $\mathcal{S}_n$'s are given as the following:

$$\mathcal{S}_m = \{i^{(m)}\},\ i^{(m)} = 1, \ldots, \ell,\ m = 1, 2, 3, 5, \tag{3.11}$$

$$\mathcal{S}_4 = \{1\}, \mathcal{S}_6 = \{0\}. \tag{3.12}$$

For phase 1 , $i^{(m)}$ keeps track of the state of queue wait time and for phases 2,3 and 5 , $i^{(m)}$ keeps track of the state of service time. Characterizing matrices $(Q, \tilde{Q}, R)$ of the GMFQ process $Z(t)$ is presented as:

$$Q = \left[\begin{array}{c|c|c|c|c|c} B & \boldsymbol{\psi} \otimes \boldsymbol{\sigma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & S - \lambda \boldsymbol{I}_\ell & \lambda \boldsymbol{I}_\ell & \boldsymbol{\nu} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & S & \mathbf{0} & \boldsymbol{\nu} \otimes \boldsymbol{\sigma} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & -\lambda & \lambda \boldsymbol{\sigma} & 0 \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & S & \boldsymbol{\nu} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} & 0 \end{array}\right], \tilde{Q} = \left[\begin{array}{c|c|c|c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \boldsymbol{\beta} & \beta_0 \boldsymbol{\sigma} & \mathbf{0} & 0 & \mathbf{0} & 0 \end{array}\right], \tag{3.13}$$

$$R = \mathbf{diag}\{\boldsymbol{I}_{4\ell+1}, -1\} \tag{3.14}$$

Table 3.2: Detailed explanation of each term in Q and $\tilde{Q}$ for $M/PH/1/2$ queue.

Term	Description
B	State of the waiting time changes from i_1 to i_2 and the system transitions from state $i_1^{(1)}$ to state $i_2^{(1)}$ with transition rate $B_{(i_1^{(1)},i_2^{(1)})}$.
$\boldsymbol{\psi} \otimes \boldsymbol{\sigma}$	Waiting time is over and the system transitions from state $i_1^{(1)}$ to state $i_2^{(2)}$ with transition rate $\boldsymbol{\psi}_{i_1^{(1)}} \boldsymbol{\sigma}_{i_2^{(2)}}$.
$S - \lambda \boldsymbol{I}_\ell$	State of the service time changes from i_1 to i_2 and the system transitions from state $i_1^{(2)}$ to state $i_2^{(2)}$. Transition rate is $S_{(i_1^{(2)},i_2^{(2)})} - \delta(i_1, i_2)\lambda$ if $i_1 = i_2$ otherwise it is $S_{(i_1^{(2)},i_2^{(2)})}$.
S	State of the service time changes from $i_1^{(m)}$ to $i_2^{(m)}$ and the system transitions from state $i_1^{(m)}$ to state $i_2^{(m)}$ with transition rate $S_{(i_1^{(m)},i_2^{(m)})}$ for $m = 3, 5$.
$\lambda \boldsymbol{I}_\ell$	A new packet has arrived and this packet is taken to the buffer and the system transitions from state $i_1^{(2)}$ to state $i_1^{(3)}$ with transition rate λ.
$\boldsymbol{\nu}$	Service time is over. If the system is in phase 2, the system transitions from state $i^{(2)}$ to state 1 with transition rate $\nu_{i^{(2)}}$. If the system is in phase 5, the system transitions from state $i^{(5)}$ to state 0 with transition rate $\nu_{i^{(5)}}$.
$\boldsymbol{\nu} \otimes \boldsymbol{\sigma}$	The service of the packet is over, the service of the packet in the buffer starts. The system transitions from state $i_1^{(3)}$ to state $i_2^{(5)}$ with a transition rate $\boldsymbol{\nu}_{i_1^{(3)}} \boldsymbol{\sigma}_{i_2^{(5)}}$.
$\lambda \boldsymbol{\sigma}$	A new packet arrived and this packet is taken to the service and the system transitions from state 1 to state $i^{(5)}$ with transition rate $\lambda \boldsymbol{\sigma}_{i^{(5)}}$.
$\boldsymbol{\beta}$	Waiting time starts and the system transitions from state 0 to state $i^{(1)}$ with transition rate $\boldsymbol{\beta}_{i^{(1)}}$.
$\beta_0 \boldsymbol{\sigma}$	Waiting time is 0 and packet is taken to the service and the system transitions from state 0 to state $i^{(1)}$ with transition rate $\beta_0 \boldsymbol{\sigma}_{i^{(1)}}$.

The fluid level $Z_f(t)$ at the beginning of phase 4 or 5 (depends on the buffer is

empty or not) represents the D_j and the fluid level just at the beginning of phase 6 represents the Φ_{j+1}. Therefore, the fluid level in the concatenation of phases 4 and 5 represents the sample cycle of AoI process and the value of fluid level $Z_f(t)$ just at the beginning of phase 6 represents a sample value of PAoI process. Since we can obtain steady-state solution of $Z(t)$, we can also obtain the exact distributions of the AoI and PAoI with the following theorem.

Theorem 4. *Let the fluid level process* $\boldsymbol{Z(t)} \sim GMFQ(Q, \tilde{Q}, R)$*, where characterizing matrices are defined in* (3.14)*. Furthermore,* $f_s(x), s \in \mathcal{S}$ *denotes the state joint pdf of the associated phase. Then, the pdfs of steady-state AoI*$(f_\Delta(x))$ *and PAoI(* $f_\Phi(x)$*) for the* $M/PH/1/2/R(r)$ *queue, can be expressed as a function of* $f_s(x)$ *in matrix exponential form by,*

$$f_\Delta(x) = \frac{\sum\limits_{s\in\mathcal{S}_4 \bigcup \mathcal{S}_5} f_s(x)}{\int\limits_0^\infty \sum\limits_{\mathcal{S}_4 \bigcup \mathcal{S}_5} f_s(x')dx'}, f_\Phi(x) = \frac{\sum\limits_{j=1}^{\ell} f_{j^{(5)}}(x)\nu_j}{\int\limits_0^\infty \sum\limits_{j=1}^{\ell} f_{j^{(5)}}(x')\nu_j dx'}. \tag{3.15}$$

Chapter 4

Multi-Source System

4.1 The General Model

To model the multi-source information update system, we consider a system model containing a number of information sources, a server, and a monitor as shown in Fig. 4.1.

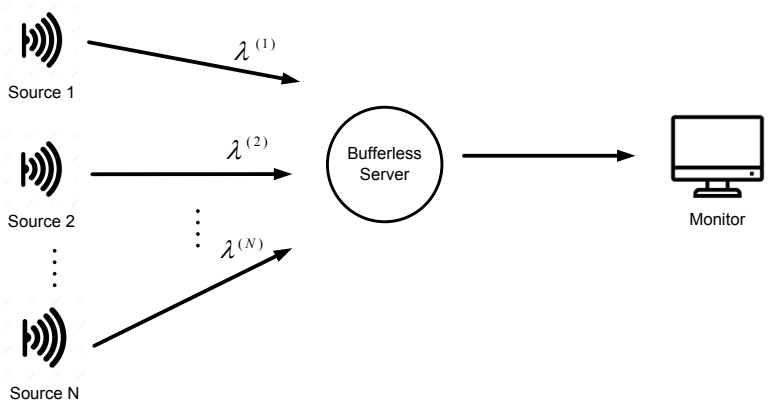

Figure 4.1: N information sources sending status update messages through a bufferless server to a remote monitor.

Similar to single-source model, information sources send their status updates

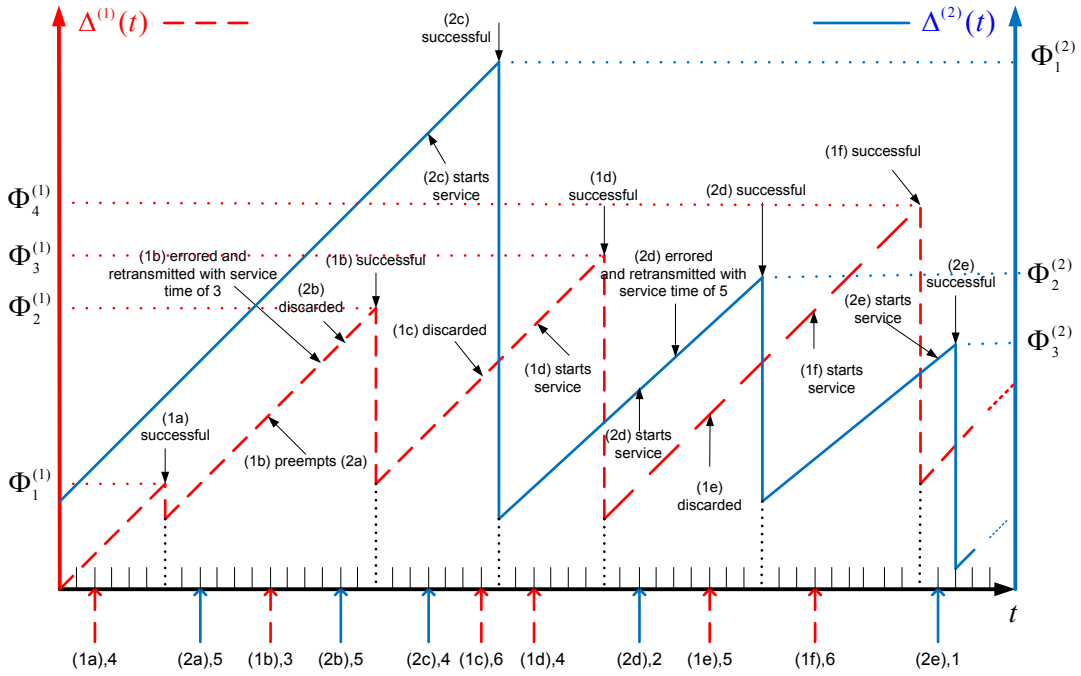

Figure 4.2: An illustration for the sample path of AoI and PAoI processes for two source and bufferless system.

to a server where the processing of these updates takes place. When the service time of an update is over, it is displayed on the monitor. In this model, a new arrival can preempt the packet in the service or errors can occur at the end of service time.

Fig. 4.2 shows a sample path of AoI and PAoI processes of two sources for bufferless, two source system. In this example, arriving packets are denoted with notation $(na), \theta$ where n indicates the packet's source, a denotes the first arriving packet and continues with $b, c, d, ...$ and θ is the service time requirement of this packet. In this example, $\Delta^{(1)}(t) = 0$ and it increases linearly until packet $1a$ successfully received at $t = 6$ and $\Delta^{(1)}(t)$ drops to the service time of this packet, then $\Phi_1^{(1)} = 6$ and $D_1^{(1)} = 4$. At $t = 12$, packet $1b$ preempts packet $2a$, and this packet is retransmitted at $t = 15$ since it is errored. At $t = 18$, packet completes the service and $\Delta^{(1)}(t)$ drops to system time of this packet, then $\Phi_2^{(1)} = 18$ and $D_2^{(1)} = 6$. Packet $1c$ is discarded when service of $2c$ continues, and when the service of $2c$ is over $\Delta^{(2)}(t)$ drops to system time of this packet, then $\Phi_1^{(2)} = 30$

and $D_1^{(2)} = 4$. At $t = 40$, the service of retransmitted source-2 packet is over, then $\Phi_2^{(2)} = 19$ and $D_2^{(2)} = 7$. And the same pattern repeats for the remaining events.

Similar to single-source model, we aim to find cdfs of AoI and PAoI processes given by Eq. 4.1 for multi-source system.

$$F_{\Delta^{(n)}}(x) = \lim_{t\to\infty} \Pr\left\{\Delta^{(n)}(t) \leq x\right\}, x \geq 0, F_{\phi(0)}(x) = \lim_{j\to\infty} \Pr\left\{\Phi_j^{(n)} \leq x\right\}, x \geq 0 \tag{4.1}$$

4.2 Bufferless Queues

In this subsection, we are interested in finding exact distributions of AoI and PAoI processes in the multi-source preemptive bufferless $M/PH/1/1$ queueing system. This queueing model can be described by Fig. 4.1 , there are N information sources, bufferless server and a remote monitor. Arrivals of information packets for each source-$n, n = 1, 2, ..., N$ are determined by the Poisson process with intensity λ_n. And total arrival intensity is defined as $\lambda = \sum_{n=1}^{N} \lambda_i$. Service time is distinct for each information source and phase-type distributed such that service time of source n is $\Theta^{(n)} \sim PH(\boldsymbol{\sigma}^{(n)}, \boldsymbol{S}^{(n)})$ with order ℓ_n, $\boldsymbol{\nu}^{(n)} = -\boldsymbol{S}^{(n)}\mathbf{1}$, and $E[\Theta^{(n)}] = \frac{1}{\mu_n}$. Total order is $l = \sum_{n=1}^{N} l_n$. Furthermore, we define preemption probability matrix P, where $P_{n,m}$ describes the probability that a new packet arrival from source m can preempt the packet in the service from source n. Three sub-cases of the preemption matrix P have been studied in the literature which are:

- $\boldsymbol{P} = \mathbf{0}$ refers to a non-preemptive system [40],
- $\boldsymbol{P} = \mathbf{1}\ \mathbf{1}^{\boldsymbol{T}}$ case is referred to as *global preemption* in [3],
- $\boldsymbol{P} = \boldsymbol{I}$ case is referred to as *self-preemption* in [4].

For each source n, the packet in the service can be preempted with intensity of traffic $\bar{\lambda}_n = \sum_{m=1}^{N} \lambda_m P_{n,m}$. At the end of the service time, a transmission error

can occur with probability e_n for each source-n packet. Moreover, we define successful transmission probability for source-n packet as $q_n = 1 - e_n$. If transmission error occurs this packet can be retransmitted or discarded with probabilities r_n and $d_n = 1 - r_n$, respectively.

We need to construct a GMFQ process $X(t)$ which consists of infinitely many cycles and these cycles begin with the arrival of first source-1 successful packet into the system and ends with successful reception of second source-1 packet. Fig. 4.3(b) describes the sample path for the fluid level process of $X_f(t)$. Each cycle consists of four phases. The first phase is constructed for the service time of the last successful packet. In this phase, packet in the service can be preempted or transmission error can occur at the end of its service time. In either case (if the packet is unsuccessful), a transition from phase 1 to phase 4 occurs. After successful completion of the service of the last packet, the system transitions from phase 1 to phase 2. In phase 2, the system is waiting for the arrival of a new packet. When the new arrival occurs, the system transitions to phase 3. If the new arrival is from source-1, successful completion of service time leads to a transition to phase 4. Otherwise (arrival is from source-n,$n = 2, ..., N$), after successful completion of service time system transitions from phase 3 to phase 2, where we are waiting for a new arrival again. In phase 4, the fluid level is brought to level 0 with drift rate -1. When the fluid level becomes 0, the system wait arbitrary time until a new cycle starts.

Each phase consists of several states that we call this set of states as $\mathcal{S}_n, n = 1, ..., 4$. The state space $\mathcal{S}$ of the modulating process $X_m(t)$ can be written as $\mathcal{S} = \bigcup_{n=1}^{4} \mathcal{S}_n$. Where $\mathcal{S}_n$'s are given as the following:

$$\begin{aligned} &\mathcal{S}_1 = \{i\},\ i = 1, \ldots, \ell_1,\ \mathcal{S}_2 = \{1\} \\ &\mathcal{S}_3 = \{(j,k)\},\ j = 1, \ldots, N,\ k = 1, \ldots, \ell_j, \end{aligned} \tag{4.2}$$

and $\mathcal{S}_4 = \{0\}$. For phase 3, j and k keep track of the source of the packet in the service and state of service time, respectively. Characterizing matrices$(Q, \tilde{Q}, R)$ of the GMFQ process $X(t)$ is presented as;

$$
Q = \left[\begin{array}{c|c|c|c|c|c|c}
\begin{array}{c}\boldsymbol{S}^{(1)} - \bar{\lambda}_1 \boldsymbol{I} \\ +e_1 r_1 \boldsymbol{V}^{(1)}\end{array} & q_1 \boldsymbol{\nu}^{(1)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \begin{array}{c}\bar{\lambda}_1 \mathbf{1} \\ +e_1 d_1 \boldsymbol{\nu}^{(1)}\end{array} \\
\hline
\mathbf{0} & -\lambda & \lambda_1 \boldsymbol{\sigma}^{(1)} & \lambda_2 \boldsymbol{\sigma}^{(2)} & \cdots & \lambda_n \boldsymbol{\sigma}^{(N)} & 0 \\
\hline
\mathbf{0} & e_1 d_1 \boldsymbol{\nu}^{(1)} & \begin{array}{c}\boldsymbol{S}^{(1)} - \bar{\lambda}_1 \boldsymbol{I} \\ +\lambda_1 P_{1,1} \boldsymbol{F}^{(1)} \\ +e_1 r_1 \boldsymbol{V}^{(1)}\end{array} & \lambda_2 P_{1,2} \boldsymbol{F}^{(2)} & \cdots & \lambda_N P_{1,N} \boldsymbol{F}^{(N)} & q_1 \boldsymbol{\nu}^{(1)} \\
\hline
\mathbf{0} & \begin{array}{c}q_2 \boldsymbol{\nu}^{(2)} \\ +e_2 d_2 \boldsymbol{\nu}^{(2)}\end{array} & \lambda_1 P_{2,1} \boldsymbol{F}^{(1)} & \begin{array}{c}\boldsymbol{S}^{(2)} - \bar{\lambda}_2 \boldsymbol{I} \\ +\lambda_2 P_{2,2} \boldsymbol{F}^{(2)} \\ +e_2 r_2 \boldsymbol{V}^{(2)}\end{array} & \cdots & \lambda_N P_{2,N} \boldsymbol{F}^{(N)} & \mathbf{0} \\
\hline
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\hline
\mathbf{0} & \begin{array}{c}q_N \boldsymbol{\nu}^{(N)} \\ +e_N d_N \boldsymbol{\nu}^{(N)}\end{array} & \lambda_1 P_{N,1} \boldsymbol{F}^{(1)} & \lambda_2 P_{N,2} \boldsymbol{F}^{(2)} & \cdots & \begin{array}{c}\boldsymbol{S}^{(N)} - \bar{\lambda}_N \boldsymbol{I} \\ +\lambda_N P_{N,N} \boldsymbol{F}^{(N)} \\ +e_N r_N \boldsymbol{V}^{(N)}\end{array} & \mathbf{0} \\
\hline
\mathbf{0}\ (\tilde{Q}_{41}) & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0\ (\tilde{Q}_{44})
\end{array}\right] \tag{4.3}
$$

where

$$
\boldsymbol{V}^{(j)} = \boldsymbol{\nu}^{(j)} \otimes \boldsymbol{\sigma}^{(j)}, \quad \boldsymbol{F}^{(j)} = \mathbf{1} \otimes \boldsymbol{\sigma}^{(j)}, \tag{4.4}
$$

and $\tilde{Q}$ is matrix of 0's except south-west block of the matrix (location is denoted by $\tilde{Q}_{41}$ in the definition of Q) is equal to σ_1 and south-east value (location is denoted by $\tilde{Q}_{44}$ in the definition of Q) is equal to -1. And the diagonal matrix of drifts is defined as $\boldsymbol{R} = \mathbf{diag}\{\boldsymbol{I}_{\ell_T+\ell_1+1}, -1\}$. Detailed description of each term in matrices $(Q, \tilde{Q})$ is given in Table 4.1.

Table 4.1: A detailed explanation of each term in Q and $\tilde{Q}$ for multi-source preemptive $M/PH/1/1$ queue.

Term	Description
$\boldsymbol{S}^{(n)} - \bar{\lambda}_n \boldsymbol{I}$	If the system is in phase 1($n = 1$), state of service time changes from i to j and system transitions from state i to j with transition rate $\boldsymbol{S}_{ij}^{(1)} - \bar{\lambda}_1 \delta(i,j)$. If the system is in phase 3, state of service time changes from state i to j and the system transitions from state (n,i) to (n,j) with transition rate $\boldsymbol{S}_{ij}^{(n)} - \bar{\lambda}_n \delta(i,j)$.
$e_n r_n \boldsymbol{V}^{(n)}$	If the system is in phase 1 ($n = 1$), the service of the source-1 packet is over, there is an error and it is retransmitted then system transitions from state i to j with transition rate $e_1 r_1 \nu_i^{(1)} \sigma_j^{(1)}$. If the system is in phase 3, service of the packet from source-n is over, there is an error and it is retransmitted then system transitions from state (n,i) to (n,j) with transition rate $e_n r_n \nu_i^{(n)} \sigma_j^{(n)}$.
$q_n \boldsymbol{\nu}^{(n)}$	If the system is in phase 1 ($n = 1$), service of the source-1 packet is over and it is successful and system transitions from state i to 1 with transition rate $q_1 \nu_i^{(1)}$. If the system is in phase 3, service of the packet from source-n is over and it is successful and system transitions from state (n,i) to 0 if $n = 1$ and (n,i) to 1 if $n = (2, ..., N)$ with transition rate $q_n \nu_i^{(n)}$.
$e_n d_n \boldsymbol{\nu}^{(n)}$	If the system is in phase 1 ($n = 1$), service of the source-1 packet is over, it is errored and discarded and system transitions from state i to 0 with transition rate $e_1 d_1 \boldsymbol{\nu}_i^{(1)}$. If the system is in phase 3, service of the packet from source-n is over, it is errored and discarded and system transitions from state (n,i) to 0 with transition rate $e_n d_n \boldsymbol{\nu}_i^{(n)}$.
$\lambda_n \boldsymbol{\sigma}^{(n)}$	A new packet arrival from source-n occurs and system transition from state 1 to (n,i) with transition rate $\lambda_n \boldsymbol{\sigma}_i^{(n)}$.
$\lambda_m P_{n,m} \boldsymbol{F}^{(m)}$	Source-n packet in the service is preempted by a new packet arrival from source-m and service is restarted for the new packet, system transitions from state (n,i) to (m,j) with transition rate $\lambda_m P_{n,m} \boldsymbol{\sigma}_j^{(m)}$.

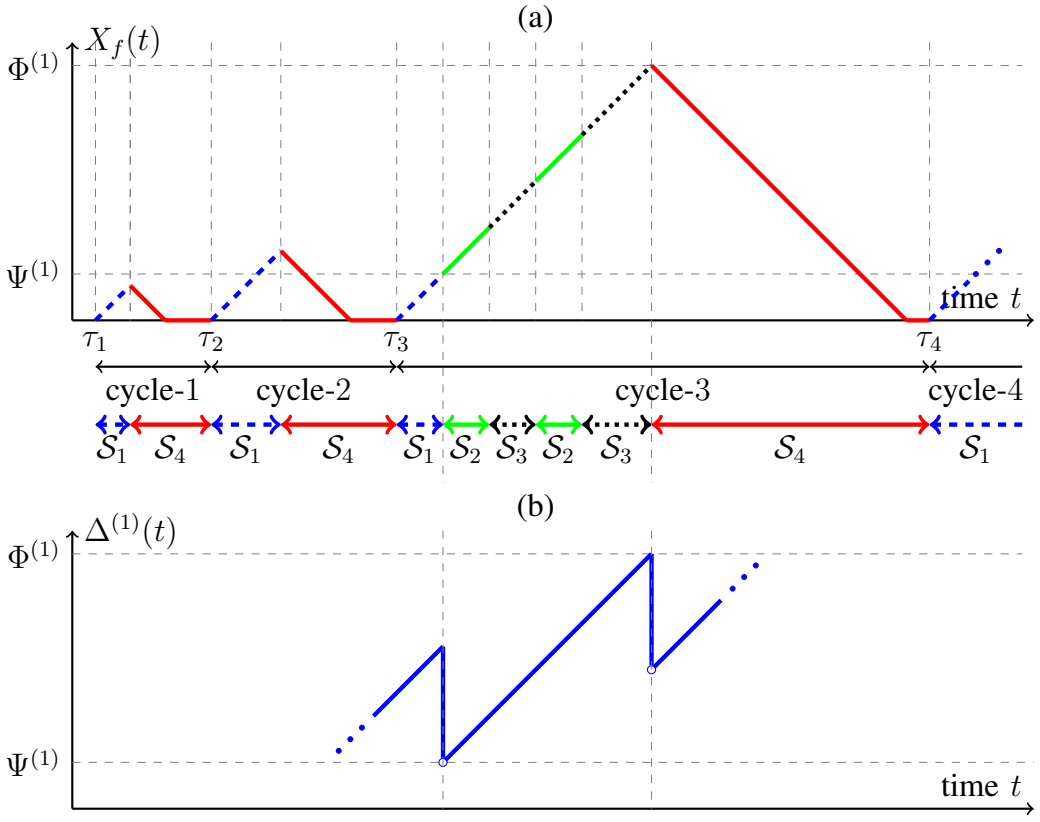

Figure 4.3: (a) Sample path of the fluid level process $X_f(t)$ (b) Sample path of the AoI process $\Delta^{(1)}(t)$. The first two cycles of $X_f(t)$ are unsuccessful cycles. Part of the successful cycle-3 spent in states of $\mathcal{S}_2$ and $\mathcal{S}_3$ overlaps with one cycle of the AoI process.

Fig. 4.3 represents the sample cycle of AoI ($\Delta^{(1)}$) process and fluid level process $X_f(t)$, respectively. After a successful reception of the first packet, fluid level reaches the value of $D^{(1)}$ which is defined as system time before. The fluid level rises until the system makes a transition to phase 4. The highest level the fluid level reaches before transition to phase 4 coincides with the PAoI value $\Phi^{(1)}$. In the Fig. 4.3 sample cycles of AoI process and fluid level process coincides starting from $D^{(1)}$ to $\Phi^{(1)}$. The fluid level during phases 2 and 3 coincides with sample cycle of AoI. Based on those observations, the exact distributions of the AoI and PAoI can be obtained with the following theorem.

Theorem 5. *Let the fluid level process* $\boldsymbol{X}(t) \sim MFQ(\boldsymbol{Q}, \tilde{\boldsymbol{Q}}, \boldsymbol{R})$ *with order* $\ell_T + \ell_1 + 2$*, where characterizing matrices are defined in* (4.3) *and* (4.2)*. Then the following closed-form expressions in the form of* (2.13) *describe the pdf of AoI,* $f_{\Delta^{(1)}}(x)$*, and PAoI,* $f_{\Phi^{(1)}}(x)$*, processes.*

$$f_{\Delta^{(1)}}(x) = \boldsymbol{g_A} e^{\boldsymbol{A}x} \boldsymbol{h_A}\ u(x),\ \boldsymbol{h_A} = \boldsymbol{H} \begin{bmatrix} \mathbf{0}_{\boldsymbol{\ell_1 \times 1}} \\ 1 \\ \mathbf{1}_{\boldsymbol{\ell_T}} \\ 0 \end{bmatrix},\ \boldsymbol{g_A} = \frac{1}{-\boldsymbol{g}\boldsymbol{A}^{-\mathbf{1}}\boldsymbol{h_A}}, \tag{4.5}$$

$$f_{\Phi^{(1)}}(x) = \boldsymbol{g_P} e^{\boldsymbol{A}x} \boldsymbol{h_P}\ u(x),\ \boldsymbol{h_P} = \boldsymbol{H} \begin{bmatrix} \mathbf{0}_{\boldsymbol{\ell_1 + 1 \times 1}} \\ q_1 \boldsymbol{\nu^{(1)}} \\ \mathbf{0}_{\boldsymbol{\ell_T - \ell_1 + 1 \times 1}} \end{bmatrix},\ \boldsymbol{g_P} = \frac{1}{-\boldsymbol{g}\boldsymbol{A}^{-\mathbf{1}}\boldsymbol{h_P}}, \tag{4.6}$$

Furthermore, by using the Lemma 1, the associated non-central moments of the AoI and PAoI processes can be obtained as the following.

$$E\left[(\Delta^{(1)})^i\right] = (-1)^{i+1}\ i!\ \boldsymbol{g_A}\boldsymbol{A}^{-\boldsymbol{(i+1)}}\boldsymbol{h_A}, i = 1, 2, \ldots, \tag{4.7}$$

$$E\left[(\Phi^{(1)})^i\right] = (-1)^{i+1}\ i!\ \boldsymbol{g_P}\boldsymbol{A}^{-\boldsymbol{(i+1)}}\boldsymbol{h_P},\ i = 1, 2, \ldots, \tag{4.8}$$

Chapter 5

Numerical Results

5.1 Single-Source System

5.1.1 Validation with Simulations

In this subsection, we will validate our proposed models with simulations for single-source system with three examples. In the first example, arrivals are determined with Poisson process with rate λ and service times are PH-type distributed with mean ρ/λ with a predetermined system load ρ and scov of the service time takes the value of c_Θ^2. To generate a PH-type distributed service time with a given c_Θ^2, following procedure is used:

- If $c_\Theta^2 = 1/j \leq 1$ for a positive integer j, we use a Erlang distribution with mean μ^{-1} and order j which is denoted as $E(\mu^{-1}, j)$.
- If $c_\Theta^2 = 1/j \leq 1$ and j is not an integer, we use a mixture of two appropriate Erlang distributions [41].
- If $c_\Theta^2 = 1/j > 1$, we use hyper-exponential distribution with balanced means which fits first two moments [41].

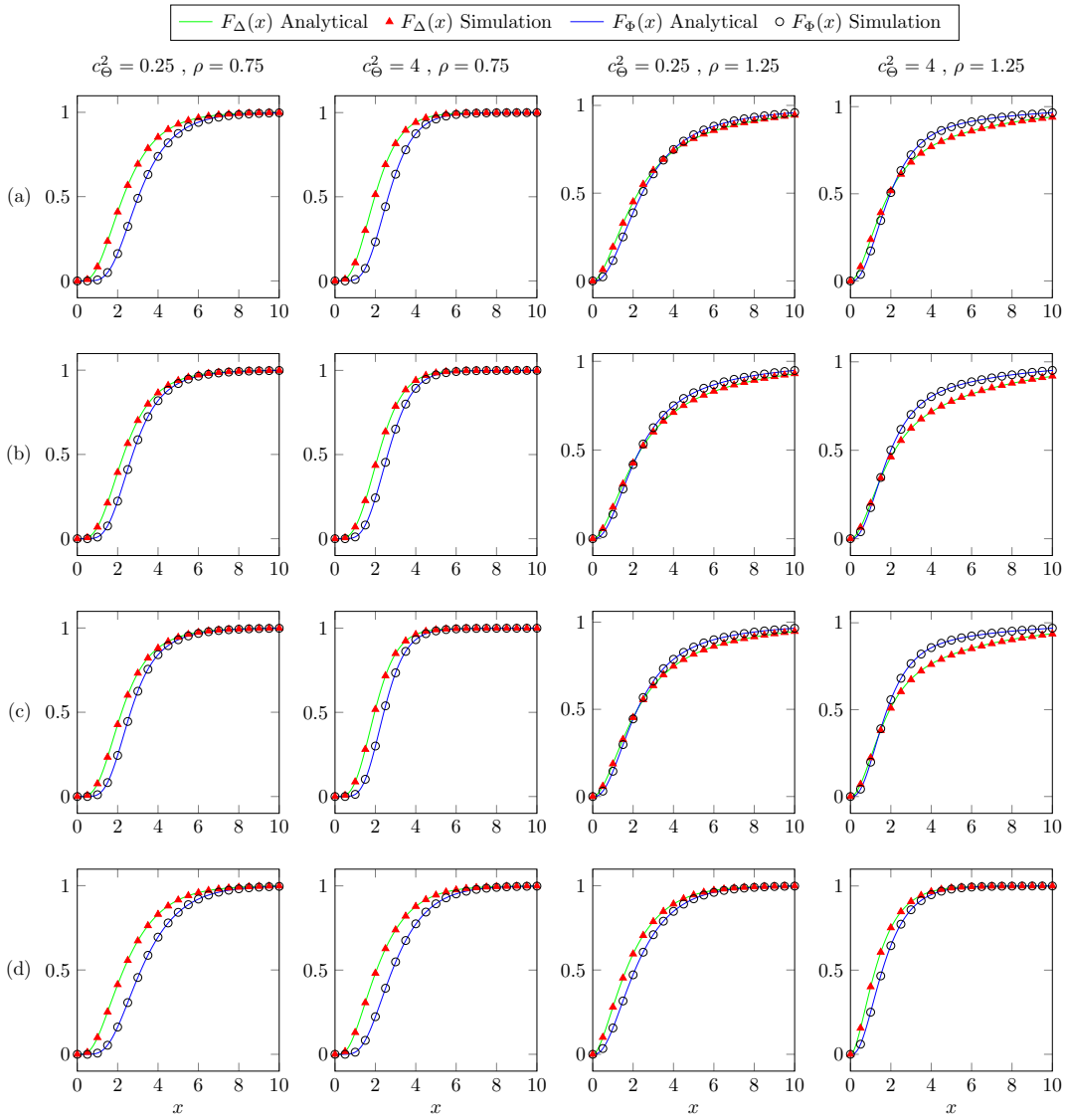

Figure 5.1: The comparison of the cdfs of the AoI and PAoI processes obtained by the proposed model and simulations for the various combination of load ($\rho \in \{0.75, 1.25\}$) and scov of service time ($c^2_\Theta \in \{0.25, 4\}$) values for (a) $M/PH/1/1$ (b) $M/PH/1/1^*$ (c) $M/PH/1/2$ (d) $M/PH/1/2^*$ queueing models.

$PH/PH/1/1/P(p)$ and $M/PH/1/1/R(r)$ queue with Poisson arrivals and for two different values of $p = (0, 1)$ and $r = (0, 1)$ are named $M/PH/1/1$, $M/PH/1/1^*$, $M/PH/1/2$ and $M/PH/1/2^*$ queue respectively. For four different pairs of scov of service time and system loads (c^2_Θ, ρ) and four different type of queues, the cdfs

of the AoI and PAoI processes obtained with the proposed model and simulations are plotted in Fig. 5.1.

For the second example, interarrival and service times are both PH-type distributed, scov of the service time is fixed to $c_\Theta^2 = 0.2$. For four different pairs of scov of interarrival time and system loads (c_Θ^2, ρ) and two different type of queueing models $PH/PH/1/1$ and $PH/PH/1/1$ the cdfs of the AoI and PAoI processes obtained with the proposed model and simulations are plotted in Fig. 5.2.

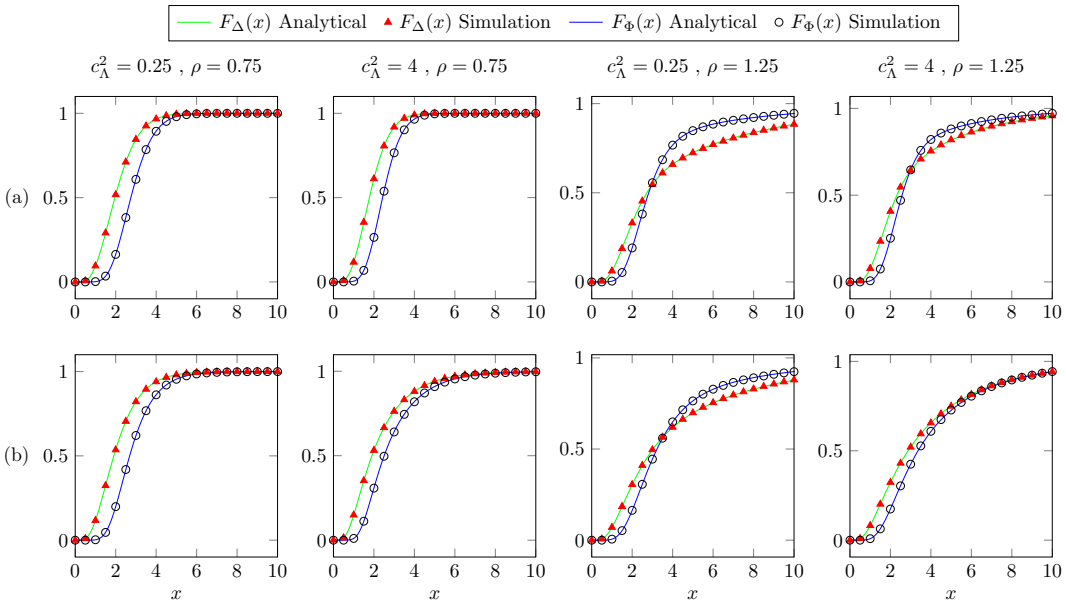

Figure 5.2: The comparison of the cdfs of the AoI and PAoI processes obtained by the proposed model and simulations for the various combination of load ($\rho \in \{0.75, 1.25\}$) and scov of arrival time($c_\Lambda^2 \in \{0.25, 4\}$) values for (a) $PH/PH/1/1$ (b) $PH/PH/1/1^*$ queueing models.

As a final example, the mean AoI ($E[\Delta]$) obtained with the proposed model and simulations with respect to packet preemption probability p and replacement probability r for $M/PH/1/1/P(p)$ and $M/PH/1/2/R(r)$ queues respectively is plotted in Fig. 5.3 for six different pairs of scov of the service time and system load (c_Θ^2, ρ). From this example, it can be inferred that optimal packet preemption probability p^* minimizing the mean AoI changes with the pair (ρ, c_Θ^2). For lower

values of c_Θ^2 and higher values of ρ (e,g., $(\rho, c_f^2) = (0.7, 2)$), increasing the packet preemption probability to 1 results in an increase in $E[\Delta]$. For the contrary case (e,g., $(\rho, c_f^2) = (0.1, 1)$) , increasing the preemption probability results in better performance. In $M/PH/1/2$ queueing model, $E[\Delta]$ is inversely proportional to packet replacement probability r for all pairs of ρ and Θ^2 we studied. Therefore $r = 1$ results in the minimum $E[\Delta]$.

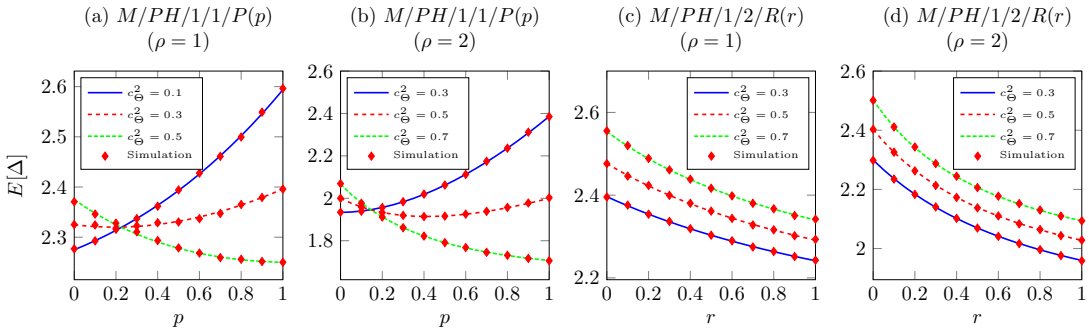

Figure 5.3: The mean AoI with respect to the packet preemption probability for the $M/PH/1/1/P(p)$ queue and with respect to packet replacement probability for the $M/PH/1/2/R(r)$ queue for three values of c_Θ^2 and two values of the ρ.

5.1.2 Validation with Existing Results

In this subsection, we validate the numerical accuracy of our proposed models with existing results in the literature for single-source system with two examples.

For the first example, closed-form expressions of $E[\Delta]$ and $E[\Delta^2]$ for preemptive LCFS $M/G/1/1^*$ and $GI/M/1/1^*$ queues and another closed-form expression of $E[\Delta]$ for $M/G/1/2^*$ are provided in [2]. Result pairs obtained with the proposed model and closed-form expressions in Ref. [2] are tabulated for three different queueing models and various cases in Table 1. The results in pairs perfectly match with each other up to 4 digits for all the scenarios studied. For the second example, the age violation probability is defined as:

$$G_\Delta(x) = 1 - F_\Delta(x).x \geq 0 \tag{5.1}$$

Table 5.1: The performance metrics $E[\Delta]$ and $E[\Delta^2]$ obtained with the closed-form expressions in Ref. [2] and the proposed method for various queueing models and their parameters.

		$E[\Delta]$	
Queueing Model		Ref. [2]	Proposed Method
$M/PH/1/1^*$	$\lambda = 0.5,\ \Theta \sim E(1,2)$	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. [2]	Proposed Method
$M/PH/1/1^*$	$\lambda = 0.5,\ \Theta \sim E(1,2)$	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

For $PH/D/1/1$ queue, where arrivals are determined with Erlang distribution, $\Lambda \sim E(\lambda^{-1}, 2)$ and the service time is deterministic with $\mu = 1$, the upper bound for $G_\Delta(x)$ is provided by [14]. To model this queue with our model, the deterministic service time is approximated by Erlang distributions $E(\mu^{-1}, j)$ with different order parameters $j = 10, 100$. With an increased value of j, we better approximate the deterministic service time since the variance of Erlang distribution j/μ^2 decreases. In Fig 9, we compared the results obtained with the proposed model, simulation, and the upper bound provided by [20]. In Fig. 5.4.(a)-(b), we plotted $G_\Delta(5)$ with respect to λ when $\mu = 1$ and $G_\Delta(x)$ with respect to x when $\lambda = 0.45$ and $\mu = 1$, respectively. According to results in Fig. 5.4 , the proposed model with approximation of service time with $E(1, 100)$ is

very close to simulation results and the upper bound is significantly loose for this example.

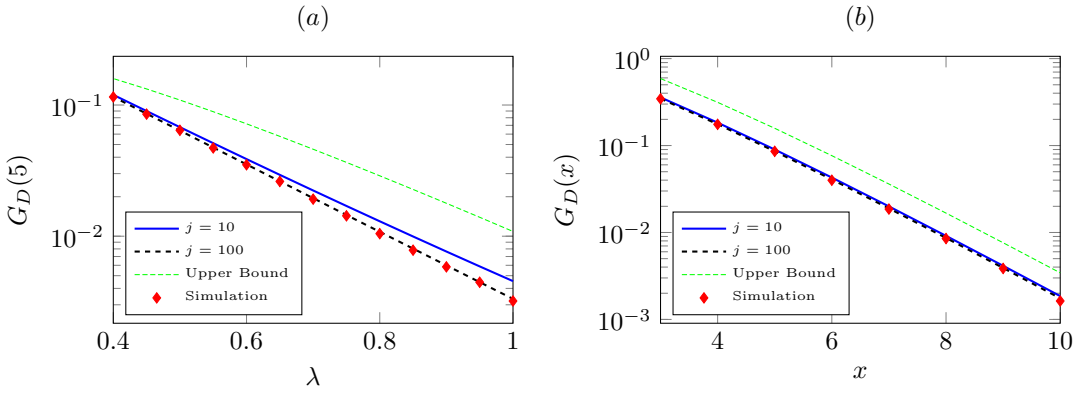

Figure 5.4: The comparison of age violation probability $G_\Delta(x)$ obtained by the proposed model, simulations and the upper bound from the proposed method by [1], as a function of arrival rate λ and the age limit x when $x = 5,\ \mu = 1$ and $\lambda = 0.45,\ \mu = 1$, respectively.

5.1.3 Analytical Results

In this subsection, we provide the results obtained with our proposed models for single-source system with three examples.

For the first example, for four different queueing models and various load parameter, the mean AoI and PAoI with respect to scov of the service time (c_Θ^2) is plotted in Fig. 5.5, and the following observations were made

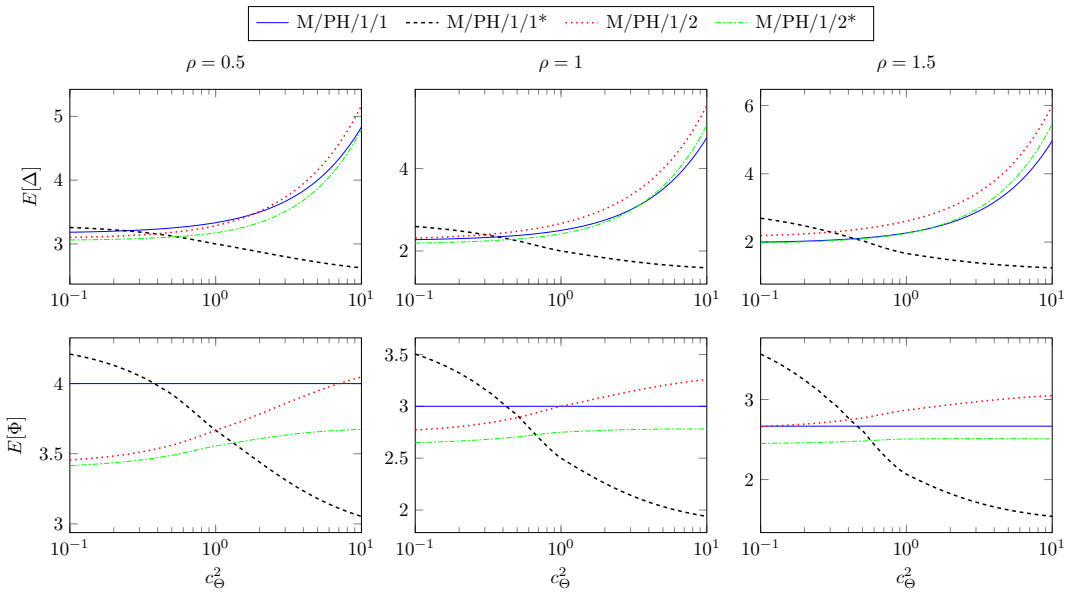

Figure 5.5: The mean AoI and the mean PAoI with respect to the scov of the service time for the four studied queueing models for various values of load ($\rho \in \{0.5, 1, 1.5.\}$)

- If there is a small variation (smaller c_Θ^2 values) in service time, interrupting the ongoing service of the packet with a new arrival with a similar service time will results in poor performance. For larger values of c_Θ^2, variations in the service time is high, preemption of packets with larger service times by the ones with shorter times is favorable. Therefore, for larger values of c_Θ^2, $M/PH/1/1^*$ is the optimal queueing model that gives the minimum $E[\Delta]$ and $E[\Theta]$, however for smaller values of c_Θ^2 $M/PH/1/1^*$ is the worst performing model.

- While $E[\Phi]$ is insensitive to changes in c_Θ^2, $E[\Delta]$ is directly proportional to c_Θ^2 when $E[\Theta]$ is fixed for $M/PH/1/1$ queue. Since $c_\Theta^2 = \sigma_\Theta^2 \mu^2$, only the second moment increases with an increase in c_Θ^2 with fixed $E[\Theta]$. $E[\Delta]$ for $M/G/1/1$ with blocking, equivalent of our $M/PH/1/1$ queue, given by [19];

$$E[\Delta] = E[\Theta]\left(\frac{\rho}{2(\rho+1)}\left(c_\Theta^2+1\right)+\frac{\rho+1}{\rho}\right) \tag{5.2}$$

  When ρ and $E[\Theta]$ is constant, $E[\Delta]$ increases with the increase in second

moment of the service time. On the other hand, PAoI for $M/PH/1/1$ system is the sum of three components which are the service time of the previous packet, time past between the completion of service of the previous packet, and the new arrival and service time of the second packet. Since no preemption is possible and arrivals are memoryless, $E[\Phi]$ can be expressed as the following;

$$E[\Phi] = 2E[\Delta] + E[\Lambda] \tag{5.3}$$

Therefore, $E[\Phi]$ is only dependent on the first moment of the service time and insensitive to changes in c_Θ^2 when $E[\Theta]$ is fixed.

- $E[\Phi]$ and $E[\Delta]$ increases with increase in c_Θ^2 for two single buffer queues $M/PH/1/2^*$ and $M/PH/1/2$. If these two models are compared, $M/PH/1/2^*$ queue always performs better than $M/PH/1/2$, since the packet in the buffer is preempted by a fresh arrival.

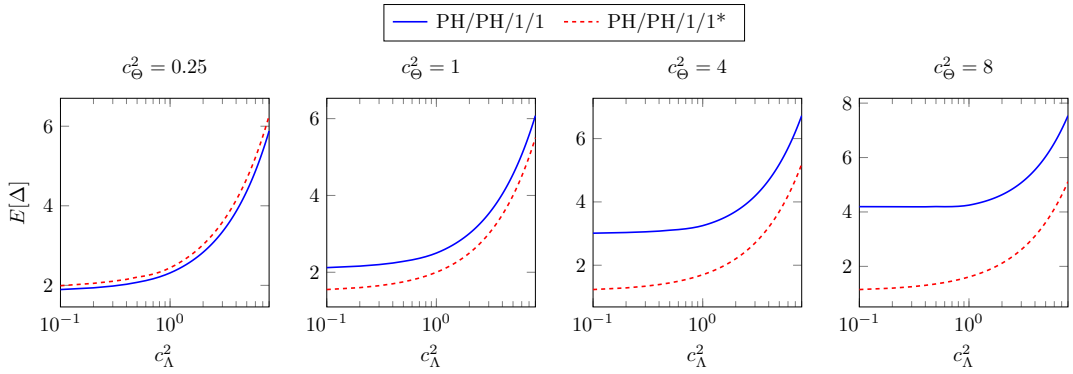

Figure 5.6: Mean AoI as a function of varying c_Λ^2 for the $PH/PH/1/1$ and $PH/PH/1/1^*$ queues for various values of scov of the service time($c_\Theta^2 \in \{0.25, 1, 4, 8\}$) and the same load ($\rho = 1$).

In the second example, for four different values of c_Θ^2 and two queueing models $M/PH/1/1$ and $M/PH/1/1^*$, the mean AoI with respect to scov of the inter-arrival time (c_Λ^2) is plotted in Fig. 5.6. It is observed that, $E[\Delta]$ increases with increased values of c_Λ^2 for both of the queues and $M/PH/1/1$ queue performs worse when $c_\Theta^2 \geq 1$.

As a final example, the queueing model (out of four) which gives rise to the minimum mean AoI is depicted in Fig. 5.7(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. 5.7(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 $M/PH/1/2^*$.

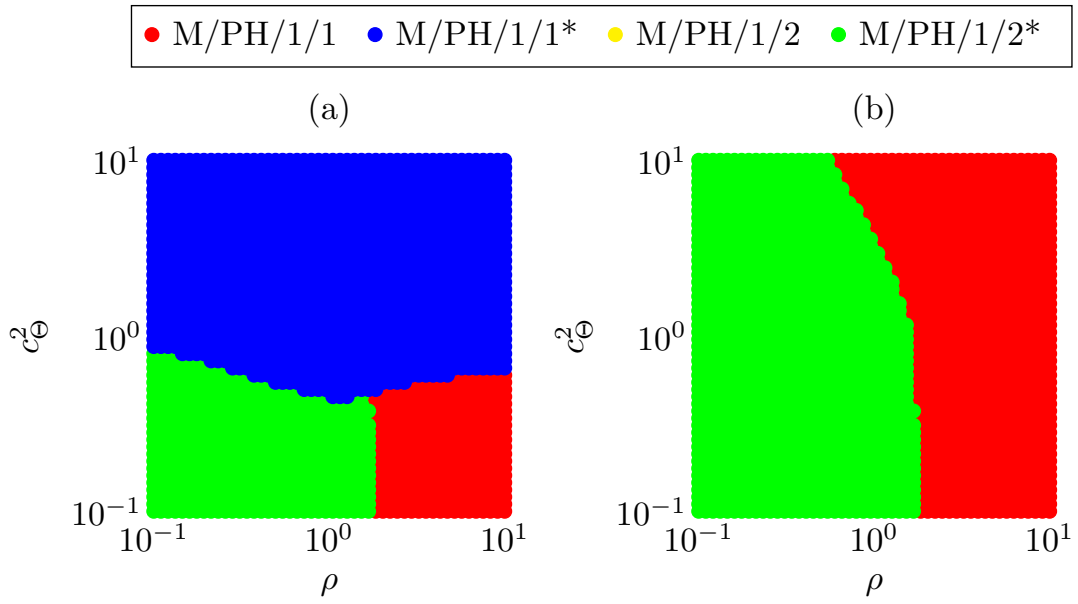

Figure 5.7: The queueing model which minimizes the mean AoI as a function of the system load ρ and scov of the service time c_Θ^2 (a) out of all four queueing models (b) out of three queueing models when $M/PH/1/1^*$ is excluded.

5.2 Multi-Source System

5.2.1 Validation with Simulations

In this subsection, we will validate our proposed models with simulations for multi-source system with two examples. We define intensity vector $\boldsymbol{\lambda}$ and it is $\boldsymbol{\lambda} = [1, 2, 3]$ for both examples. Apart from global preemption and self preemption, we also include prioritized preemption. With prioritized preemption policy, source-m packet is allowed to preempt the source-n packet in service if only if $m \leq n$, therefore source-1(3) has the highest(lowest) priority. Preemption matrix P for this policy is a lower triangular matrix where all entries below and at the main diagonal are 1. The cdfs of the AoI and PAoI processes of all sources obtained with the proposed model and simulations are plotted in Fig. 5.8 and Fig. 5.9 for $c_\Theta^2 = 0.25$ and $c_\Theta^2 = 0.5$ respectively. From these examples, we observe that cdf curves of all sources in Fig. 5.8 are steepier than in Fig. 5.9. It implies that preemption is favorable in terms of AoI with increased values of c_Θ^2 as in the single-source case. Furthermore, the cdf curves of AoI and PAoI of all the three sources are very close to each other with prioritized preemption compared to other policies since the highest priority is given to source with the lowest intensity and vice versa while sources have no priority with the other two preemption policies.

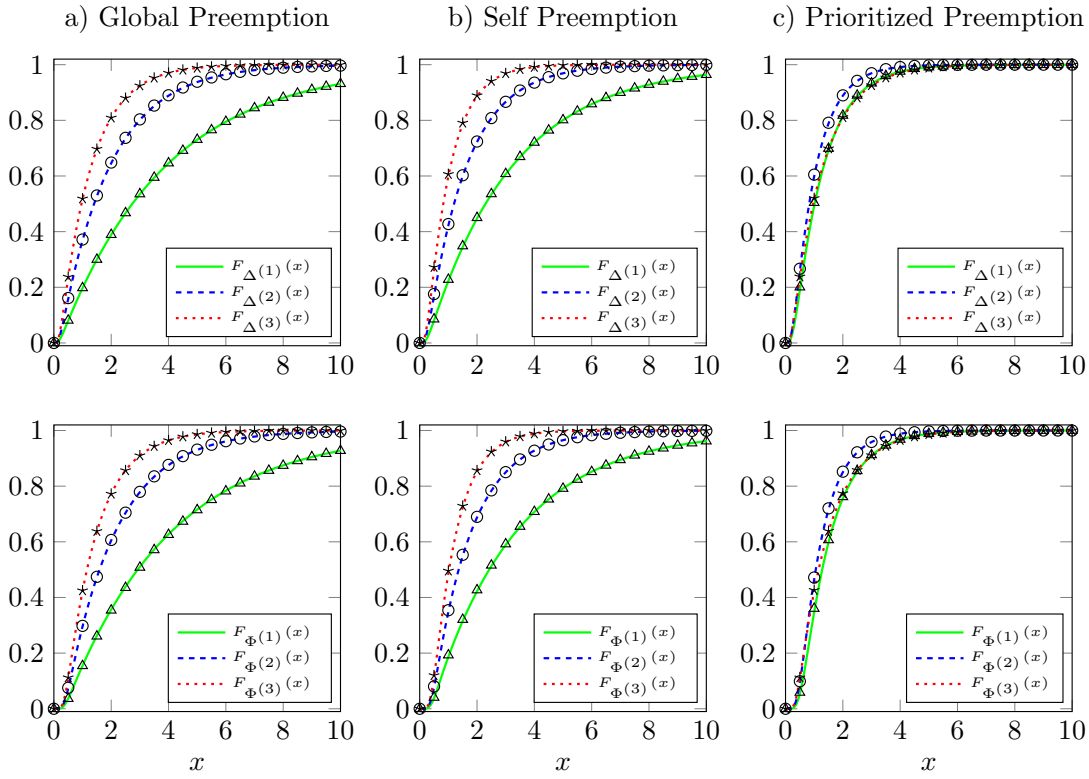

Figure 5.8: The comparison of the cdfs of the AoI and PAoI processes of a 3 source system obtained by the proposed model and simulations for three different preemption policies (a) global preemption, (b) self preemption, (c) prioritized preemption, when $\boldsymbol{\lambda} = [1, 2, 3]$, $\rho = 2/3$, $c_{\Theta}^2 = 1/4$ and $r_i = 1, e_i = 0.1,\ i = 1, 2, 3$.

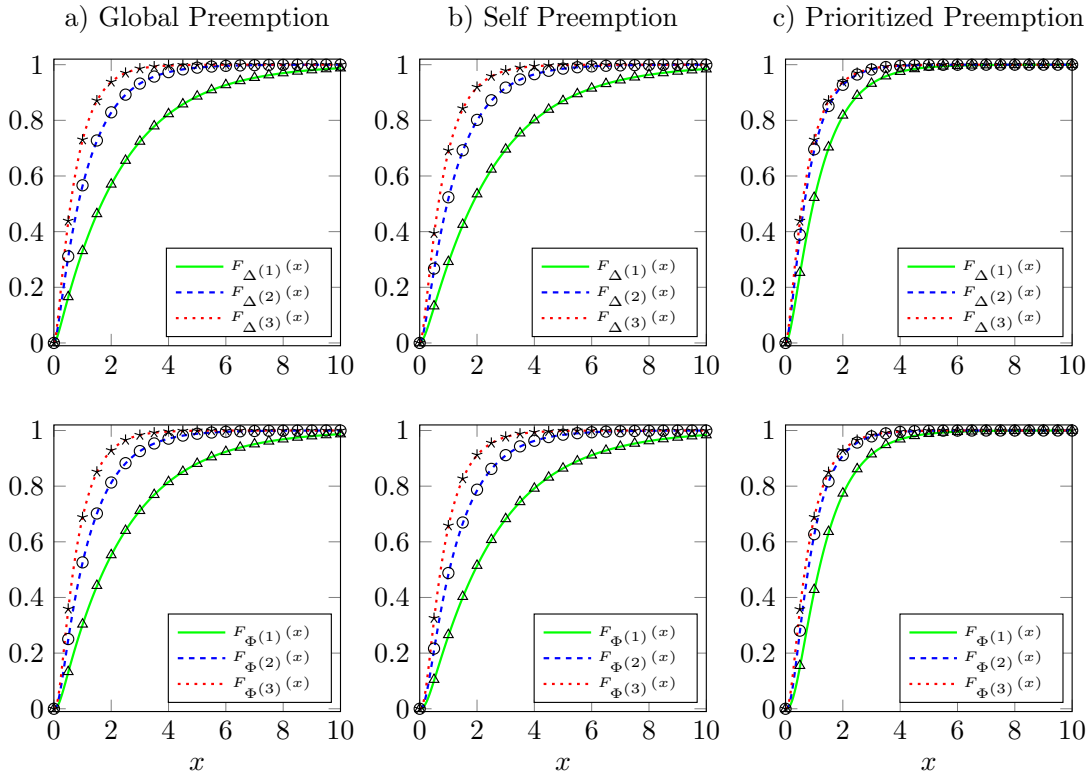

Figure 5.9: The comparison of the cdfs of the AoI and PAoI processes of a 3 source system obtained by the proposed model and simulations for three different preemption policies (a) global preemption, (b) self preemption, (c) prioritized preemption, when $\boldsymbol{\lambda} = [1, 2, 3]$, $\rho = 2/3$, $c_{\Theta}^2 = 1/2$ and $r_i = 1, e_i = 0.1,\ i = 1, 2, 3$.

5.2.2 Validation with Existing Results

In this subsection, we validate the numerical accuracy of our proposed models with existing results in the literature for multi-source system with two examples.

For the first example of multi-source case, closed-form expression for mean AoI for M/M/1 LCFS queue with global preemption, memoryless homogenous service time for each source and no transmission error is derived by [3]. Results pair for mean AoI obtained with the proposed model and closed-form expressions in Ref. [3] are tabulated for various values of the load parameter and two different traffic intensity vectors in Table 5.2. The results in pairs match perfectly up to 4 digits

for all scenarios studied.

Table 5.2: Mean AoI for each of the three sources obtained with the closed-form expressions in Ref. [3] and the proposed method for various values of the load parameter ρ. Case A: $\boldsymbol{\lambda} = [1, 2, 3]$, Case B: $\boldsymbol{\lambda} = [1, 4, 16]$.

		$E[\Delta^{(1)}]$		$E[\Delta^{(2)}]$		$E[\Delta^{(3)}]$	
Case	ρ	[3]	Prop.	[3]	Prop.	[3]	Prop.
A	0.50	1.5000	1.5000	0.7500	0.7500	0.5000	0.5000
	0.75	1.7500	1.7500	0.8750	0.8750	0.5833	0.5833
	1.00	2.0000	2.0000	1.0000	1.0000	0.6667	0.6667
	1.25	2.2500	2.2500	1.1250	1.1250	0.7500	0.7500
	1.50	2.5000	2.5000	1.2500	1.2500	0.8333	0.8333
B	0.50	1.5000	1.5000	0.3750	0.3750	0.0938	0.0938
	0.75	1.7500	1.7500	0.4375	0.4375	0.1094	0.1094
	1.00	2.0000	2.0000	0.5000	0.5000	0.1250	0.1250
	1.25	2.2500	2.2500	0.5625	0.5625	0.1406	0.1406
	1.50	2.5000	2.5000	0.6250	0.6250	0.1563	0.1563

Furthermore, a closed-form expression for mean AoI for M/M/1 LCFS queue with self preemption, memoryless homogeneous service time each source with transmission error is derived by [4]. Results pair for mean AoI obtained with the proposed model and closed-form expressions in Ref.[4] are tabulated for various values of the load ρ and transmission error parameter and two different traffic intensity vectors in Table 5.3. Obtained results match with each other up to 4 digits with the exception of one pair where results match up to 3 digits.

Table 5.3: Mean AoI for each of the three sources obtained with the closed-form expressions in Ref. [4] and the proposed method for various values of the load ρ and the error parameter e. Case A: $\boldsymbol{\lambda} = [1, 2, 3]$, Case B: $\boldsymbol{\lambda} = [1, 4, 16]$.

			$E[\Delta^{(1)}]$		$E[\Delta^{(2)}]$		$E[\Delta^{(3)}]$	
Case	ρ	e	[4]	Prop.	[4]	Prop.	[4]	Prop.
A	0.5	0.04	1.5839	1.5839	0.7971	0.7971	0.5319	0.5319
		0.10	1.6880	1.6880	0.8492	0.8492	0.5667	0.5667
		0.25	2.0214	2.0214	1.0159	1.0159	0.6778	0.6778
	1.0	0.04	2.1429	2.1429	1.0833	1.0833	0.7222	0.7222
		0.10	2.2817	2.2817	1.1528	1.1528	0.7685	0.7685
		0.25	2.7262	2.7262	1.3750	1.3750	0.9167	0.9167
	1.5	0.04	2.7042	2.7042	1.3688	1.3687	0.9109	0.9109
		0.10	2.8778	2.8778	1.4556	1.4556	0.9688	0.9688
		0.25	3.4333	3.4333	1.7333	1.7333	1.1540	1.1540
B	0.5	0.04	1.5699	1.5699	0.3965	0.3965	0.0990	0.0990
		0.10	1.6740	1.6740	0.4225	0.4225	0.1055	0.1055
		0.25	2.0074	2.0074	0.5059	0.5059	0.1264	0.1264
	1.0	0.04	2.1050	2.1050	0.5370	0.5370	0.1334	0.1334
		0.10	2.2439	2.2439	0.5717	0.5717	0.1421	0.1421
		0.25	2.6883	2.6883	0.6829	0.6829	0.1699	0.1699
	1.5	0.04	2.6423	2.6423	0.6780	0.6780	0.1675	0.1675
		0.10	2.8159	2.8159	0.7214	0.7214	0.1784	0.1784
		0.25	3.3714	3.3714	0.8603	0.8603	0.2131	0.2131

5.2.3 Analytical Results

In this subsection, we provide the results obtained with our proposed models for multi-source system with four examples.

We work on a system with two sources and homogeneous service times for the

first example. We firstly define system cost function $C(\alpha)$ as:

$$C(\alpha) = E[\Delta^{(1)}] + \alpha E[\Delta^{(2)}],\ 0 \leq \alpha \leq 1, \tag{5.4}$$

In this function, α represents the relative importance of mean AoI of source-2 to source-1. When $\alpha = 1$ two sources are equally important and for the opposite case ($\alpha = 0$) the mean age of source-2 is insignificant. For the following two examples, we will optimize the entries of preemption probability matrix P. We define P such as:

$$P = \begin{bmatrix} P_d & P_{2,1} \\ P_{1,2} & P_d \end{bmatrix} \tag{5.5}$$

Using our proposed method, we perform brute force optimization on P_d, $P_{2,1}$, $P_{1,2}$ to find their optimal values to obtain minimum value of $C(\alpha)$. The resolution of parameters for optimization is chosen as 0.02. For various values of traffic intensity vector, c_Θ^2, and α values, optimal values of entries are tabulated in Table 5.4. The following observations were made;

- The entry represents the self preemption probability P_d^* depends only on the value of c_Θ^2. It is always 1, when $c_\Theta^2 = 1$, and 0 otherwise.
- For increasing values of α, importance of mean AoI of source-2 increases. As a result, the value of $P_{1,2}*$ increases and $P_{2,1}*$ decreases.

Table 5.4: Optimum preemption parameters P_d^*, $P_{1,2}^*$, and $P_{2,1}^*$, which minimize $C(\alpha)$ for various values of the traffic intensity vector $\boldsymbol{\lambda}$, c_Θ^2, and α when $\rho = 1$ and $r_i = 0.9,\ e_i = 0.1,\ i = 1, 2$

c_Θ^2	α	$\boldsymbol{\lambda}$								
		$[1,2]$			$[1,1]$			$[2,1]$		
		P_d^*	$P_{2,1}^*$	$P_{1,2}^*$	P_d^*	$P_{2,1}^*$	$P_{1,2}^*$	P_d^*	$P_{2,1}^*$	$P_{1,2}^*$
1/16	.25	0	1	0	0	.68	0	0	0	0
	.50	0	1	0	0	.28	0	0	0	.26
	.75	0	.90	0	0	.08	0	0	0	.50
	1	0	.68	0	0	0	0	0	0	.68
1/4	.25	0	1	0	0	.72	0	0	0	0
	.50	0	1	0	0	.32	0	0	0	.30
	.75	0	.96	0	0	.10	0	0	0	.54
	1	0	.74	0	0	0	0	0	0	.74
1	.25	1	1	0	1	1	.12	1	1	1
	.50	1	1	0	1	1	.54	1	.58	1
	.75	1	1	.04	1	1	.86	1	.32	1
	1	1	1	.18	1	1	1	1	.18	1

As a second example, for three different values of α and two different pair of traffic intensities (ρ_1, ρ_2), cost function $C(\alpha)$ is plotted with respect to load parameter for four preemption policies which are global preemption, self preemption, no preemption and optimum preemption in Fig. 5.10. For optimum preemption policy, entries of preemption probability matrix P are optimized with brute force optimization. For this example, $c_\Theta^2 = 1$ since service time is exponentially distributed. We observe that optimum preemption outperforms all other policies except one instance ($\alpha = 1$ and $\rho_1 = \rho_2$) where it performs the same with global preemption. If traffic intensities and importance of sources are uneven, improvement from optimum preemption increases. We also observe that global preemption outperforms self and no preemption policies in our example.

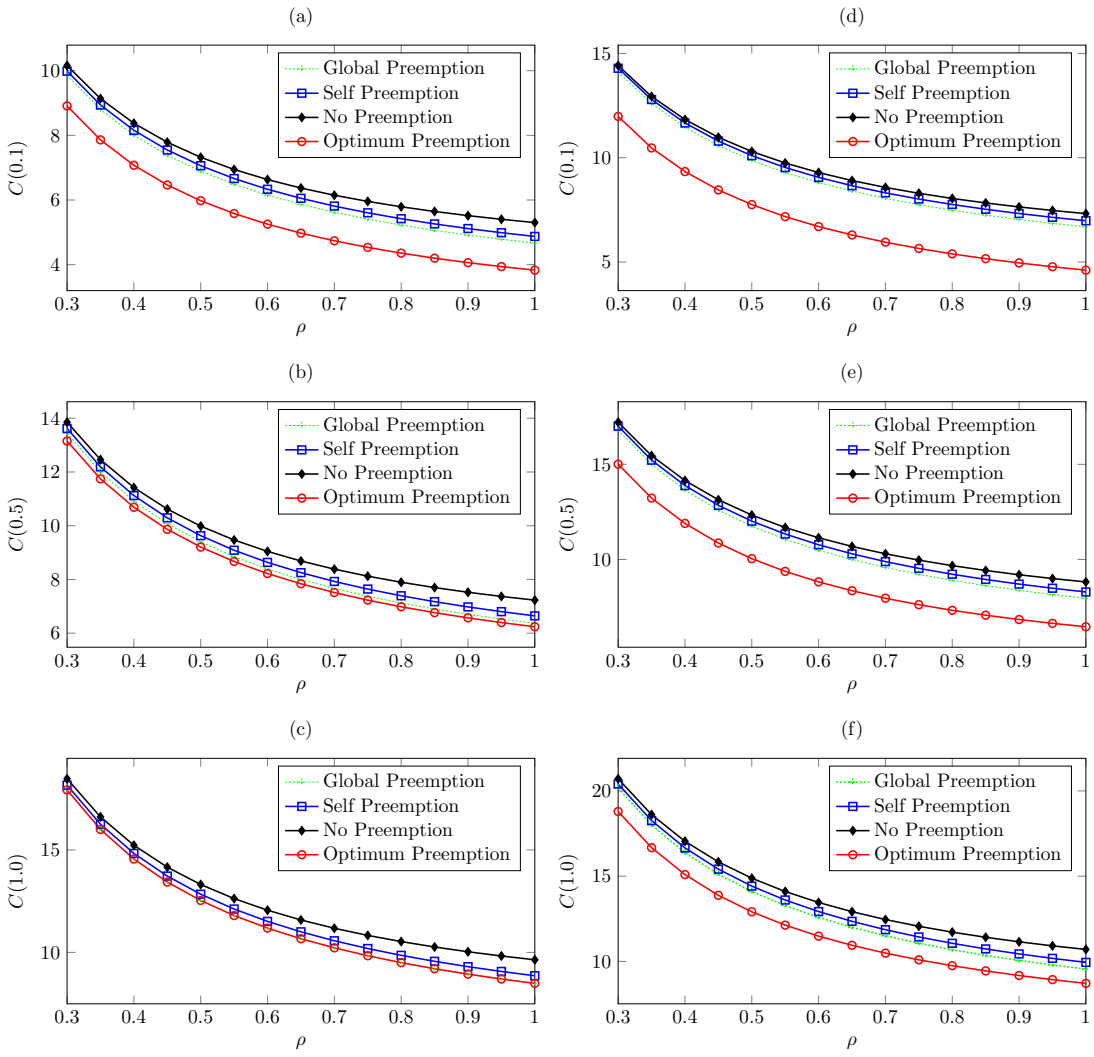

Figure 5.10: The cost function $C(\alpha)$ as a function of the ρ for four preemption policies and various values of the relative importance($\alpha \in \{0.1, 0.5, 1\}$) when $\mu = 1$, $r = 0.9$ and $e = 0.1$. For the first three examples (a)-(c), arrival rates of sources is the same ($\rho_2 = 2\rho_1$), and the arrival rate of the second source is as twice as the first source ($\rho_2 = 2\rho_1$) for the other examples (d)-(f).

For this example, we study a two-source M/M/1/1 system with homogeneous service requirements. Particularly, we set $\mu_n = \mu = 1,\ n = 1, 2$. The packet error probability, retransmission probability, and the traffic intensity are also assumed to be independent of the source index, i.e., $e_n = e, r_n = r, \lambda_n = \lambda, n = 1, 2$. We

depict in Fig. 5.11 the mean system AoI $E[\Delta]$ which is defined as the arithmetic average of the mean AoI values of the two sources, as a function of the error parameter e, for three values of the parameter r, and for two choices of the load parameter ρ when NP and GP are deployed at the server. We observe that the mean system AoI increases with increased error probability as expected for both NP and GP policies. With GP, for a given error probability e, the mean system AoI appears to be larger for smaller retransmission probability r. For GP, we conclude that the best system AoI performance is achieved with the choice of $r = 1$ with this observation being more apparent for smaller loads and for moderately lossy communications channels. The impact of r is more obscured with NP at relatively higher loads since a packet which is errored is already stale since it has been generated at least one service time back. If this packet is retransmitted, it may block a potential fresh information packet arrival within its service time (more likely in high load scenarios). However, if discarded, in the case of lack of a fresh information packet arrival for an extended time (more likely in low load scenarios), the age of the source would increase. We tend to believe that the choice of $r = 1$ would give rise to acceptable performance for relatively low loads for NP but at higher loads, further work needs to be done to derive the optimal retransmission probabilities.

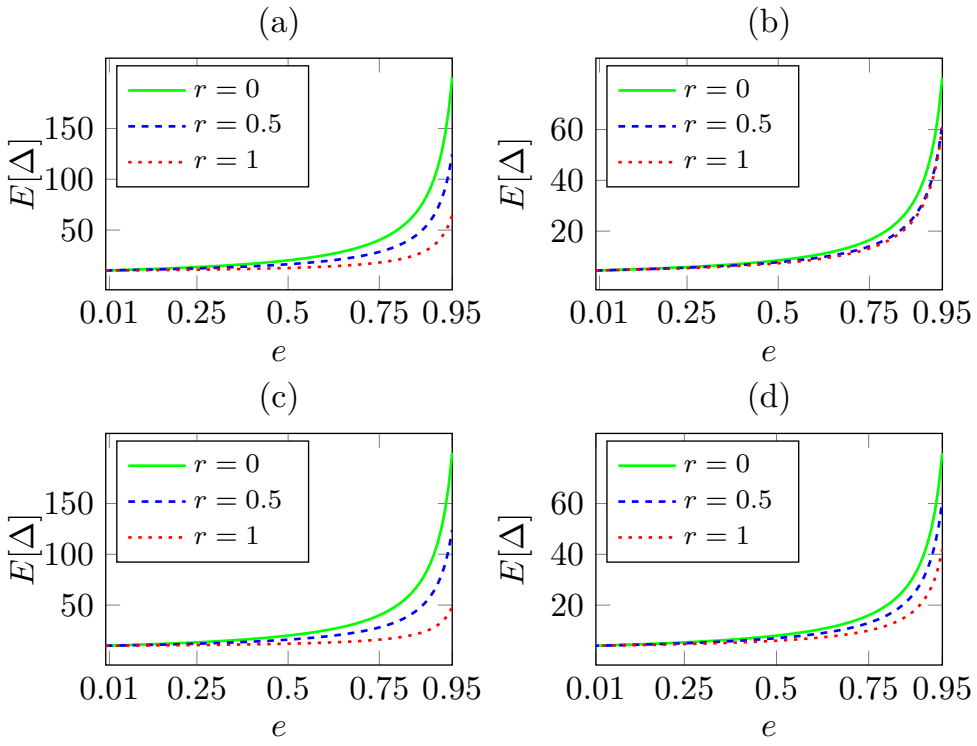

Figure 5.11: The mean system AoI $E[\Delta]$ as a function of the error probability e for three values of r (a) NP ($\rho = 0.25$) (b) NP ($\rho = 1$) (c) GP ($\rho = 0.25$) (d) GP ($\rho = 1$)

In this example, we study the same homogeneous system as in the previous example but with PH-distributed service time requirements and for general $N \geq 2$. In particular, we set the homogeneous error and retransmission probabilities as $e = 0.1, r = 1$. For service times, we employ a PH-type distribution with mean $1/\mu$ and squared coefficient of variation (scov) $c_\Theta^2 = \sigma_\Theta^2 \mu^2$ where σ_Θ^2 denotes the variance of the service time. We also fix $\lambda = \mu/N$ so as to achieve a unit system load. For service times, we employ a PH-type distribution with mean $1/\mu$ and squared coefficient of variation (scov) $c_\Theta^2 = \sigma_\Theta^2 \mu^2$ where σ_Θ^2 denotes the variance of the service time. For the purpose of constructing the PH-type distribution with given mean and scov, we used the same procedure presented in single-source section. The mean residual service time of the positive random variable Θ is denoted by $m_r(x) = E[\Theta - x \mid \Theta > x]$ [42]. The Erlang distribution with $c_\Theta^2 < 1$ satisfies $m_r(x) < m_r(0) = E[\Theta]$ whereas the H_2 distribution with $c_\Theta^2 > 1$

ensures $m_r(x) > E[\Theta]$ [42]. For the exponential distribution, it is clear that $m_r(x) = E[\Theta]$ for all x due to the memoryless property. Therefore, when $c_\Theta^2 \geq 1$ in our numerical examples, a fresh packet is not only more timely than the packet in service but it also does not have a longer residual service time in expectation, favoring preemption in this regime. Similarly, when $c_\Theta^2 < 1$ in our numerical examples, a fresh packet is more timely than the packet in service but it has a relatively longer residual service time in expectation which gives rise to a trade-off in this regime. To study the effect of c_Θ^2 in more detail, we fix $\mu = 1$, and depict the mean system AoI $E[\Delta]$ as a function of c_Θ^2 using three different preemption policies NP, GP, and SP, for three values of $N \in \{2, 4, 8\}$ in Fig. 5.12 which reveals that the mean system AoI for GP and SP decreases whereas the performance of NP substantially deteriorates with increased c_Θ^2. Moreover, there appears to be a threshold value for c_Θ^2 above which GP outperforms NP with this threshold slightly increasing with increased N but always staying below unity. Similar thresholds appear to exist above which SP outperforms NP or GP outperforms SP and all these threshold values are not necessarily identical but quite close to each other.

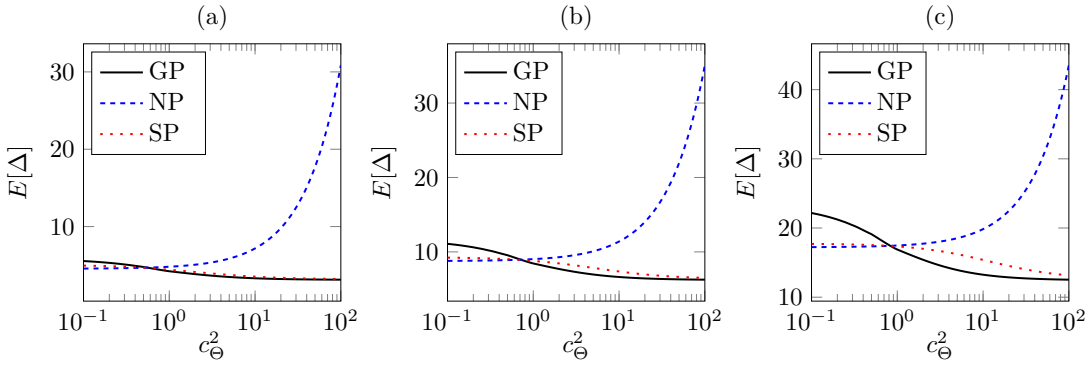

Figure 5.12: The mean system AoI $E[\Delta]$ as a function of the scov parameter c_Θ^2 (a) $N = 2$ (b) $N = 4$ (c) $N = 8$

Finally, we study the optimum choice of traffic intensities for a two-source M/M/1/1 system deploying an a-priori given preemption policy with homogeneous service times. We seek to minimize the total packet generation rate $\lambda_T = \lambda_1 + \lambda_2$ that is indicative of the power needed for sensing while satisfying

the following PAoI violation probabilities:

$$\Pr\{\Phi^{(1)} > D\} < \varepsilon,\ \Pr\{\Phi^{(2)} > D\} < \varepsilon\alpha^{-1},\ \varepsilon \leq \alpha \leq 1, \tag{5.6}$$

for a violation threshold parameter D, violation tolerance parameter ε and α being a parameter representing the relative importance of the second source with respect to the first; when $\alpha \to \varepsilon$, we do not have a violation constraint at all for source-2 whereas when $\alpha = 1$, source-2 violation constraint is as strict as that of source-1. As an example, we set $\mu = 1$, $D = 10$, $\varepsilon = 10^{-2}$ and for a given value of α, we use the analytical model and exhaustive search to find the optimum source-n traffic intensities denoted by $\lambda_n^*, n = 1, 2$ which are depicted in Fig. 5.13 as a function of the parameter α. As we relax the violation constraint for source-2, it is possible to substantially reduce the individual traffic intensities. Moreover, as expected, the ratio λ_1^*/λ_2^* increases with decreased α.

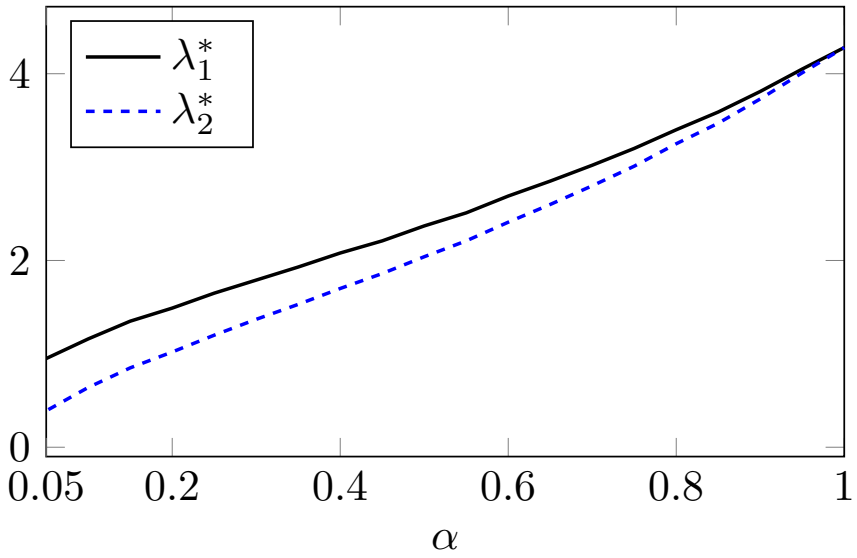

Figure 5.13: The optimum source-n traffic intensities $\lambda_n^*, n = 1, 2$ with respect to the parameter $\alpha \in [0.05, 1]$.

Chapter 6

Conclusions

In this thesis, we focus on the AoI and PAoI distributions for several well-known queueing models to study the effects of probabilistic preemption and probabilistic replacement among packets, and also distributions of service and interarrival times on the AoI performance. We obtained the probability density function of AoI and PAoI in matrix exponential form for general queueing systems. After deriving analytical expressions via numerical methods, we have validated our results with simulations and also with existing results in the literature.

For the single-source model, the optimum preemption probability is found to be dependent on the traffic intensity mix as well as the squared coefficient of variation (scov) of service times. For lower values of scov, increasing packet preemption probability resulted in worse performance, i.e., higher average AoI. For the contrary case where scov is higher, increasing the probability of preemption is found to be useful for AoI-related performance. We observed that, in some cases, probabilistic preemption with a preemption probability choice being between 0 and 1, performs better than the choice of probabilities at the boundary ($p = 0$ (no-preemption) and $p = 1$ (always preempt)). Furthermore, for deterministic distribution of service times, the upper bound from existing results is shown to be very loose compared to simulations. We also identified regions in the parameter space where the best buffer-management scheme is shown.

For the multi-source case, similarly, we have shown that preemption is also effective when scov is high. For the two-source system, we were able to show that probabilistic preemption essentially outperforms no-preemption or global-preemption. Moreover, the performance gain from optimization of preemption probabilities significantly increases when imparity in traffic intensities and relative importance of sources become more apparent.

Finding optimum preemption probabilities for the multi-source case with more than two classes by means of machine learning is left for future work. Also, the analysis of AoI in the multi-source M/PH/1/2 queueing model can be done by means of Markov fluid queues, where not only the effects of preemption but the replacement probabilities among classes are also crucial for performance. Moreover, with more detailed analysis, a more powerful buffer-management scheme that is adaptive to the changes in scov, the relative importance of sources, and distributions of service times could be proposed.

Bibliography

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

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

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

[4] S. Farazi, A. G. Klein, and D. Richard Brown, "Average age of information in multi-source self-preemptive status update systems with packet delivery errors," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 396–400, 2019.

[5] M. G. Rodriguez, L. E. Ortiz Uriarte, Yi Jia, K. Yoshii, R. Ross, and P. H. Beckman, "Wireless sensor network for data-center environmental monitoring," in *2011 Fifth International Conference on Sensing Technology*, pp. 533–537, 2011.

[6] W. Liang, Z. Li, H. Zhang, S. Wang, and R. Bie, "Vehicular ad hoc networks: architectures, research issues, methodologies, challenges, and trends," *International Journal of Distributed Sensor Networks*, vol. 11, no. 8, p. 745303, 2015.

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

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

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

[10] N. Akar and O. Dogan, "Discrete-time queueing model of age of information with multiple information sources," *arXiv preprint arXiv:2007.11650*, 2020. (Accepted for publication in IEEE Internet of Things Journal).

[11] 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, 2020.

[12] O. Dogan and N. Akar, "The multi-source preemptive M/PH/1/1 queue with packet errors: Exact distribution of the age of information and its peak," *arXiv preprint arXiv:2007.11656*, 2020.

[13] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 576–580, IEEE, 2017.

[14] J. P. Champati, H. Al-Zubaidy, and J. Gross, "Statistical guarantee optimization for AoI in single-hop and two-hop FCFS systems with periodic arrivals," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 365–381, 2021.

[15] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 486–493, IEEE, 2017.

[16] A. Javani, M. Zorgui, and Z. Wang, "Age of information in multiple sensing," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2019.

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

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

[19] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 131–135, 2017.

[20] E. Najm and R. Nasser, "Age of information: The gamma awakening," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2574–2578, Ieee, 2016.

[21] S. K. Kaul and R. D. Yates, "Timely updates by multiple sources: The M/M/1 queue revisited," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, IEEE, 2020.

[22] R. D. Yates, "Age of information in a network of preemptive servers," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 118–123, IEEE, 2018.

[23] C. Kam, J. P. Molnar, and S. Kompella, "Age of information for queues in tandem," in *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pp. 1–6, IEEE, 2018.

[24] M. Moltafet, M. Leinonen, and M. Codreanu, "On the age of information in multi-source queueing models," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5003–5017, 2020.

[25] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6419–6428, 2018.

[26] K. Chen and L. Huang, "Age-of-information in the presence of error," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2579–2583, IEEE, 2016.

[27] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 15–28, 2019.

[28] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Non-linear age of information in a discrete time queue: Stationary distribution and average performance analysis," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.

[29] V. Tripathi, R. Talak, and E. Modiano, "Age of information for discrete time queues," *arXiv preprint arXiv:1901.10463*, 2019.

[30] H. H. Yang, A. Arafa, T. S. Quek, and H. Poor, "Optimizing information freshness in wireless networks: A stochastic geometry approach," *IEEE Transactions on Mobile Computing*, vol. 20, pp. 2269–2280, jun 2021.

[31] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the EM algorithm," *Scandinavian Journal of Statistics*, pp. 419–441, 1996.

[32] S. Asmussen and C. A. O'cinneide, "Matrix-exponential distributions," *Encyclopedia of Statistical Sciences*, 2004.

[33] Q.-M. He, *Fundamentals of Matrix-Analytic Methods*. Springer, 2013.

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

[35] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, third ed., 1996.

[36] N. Akar and K. Sohraby, "System-theoretical algorithmic solution to waiting times in semi-Markov queues," *Perform. Eval.*, vol. 66, p. 587–606, Nov. 2009.

[37] V. G. Kulkarni, "Fluid models for single buffer systems," in *Frontiers in Queueing: Models and Applications in Science and Engineering* (J. H. Dshalalow, ed.), ch. Fluid Models for Single Buffer Systems, pp. 321–338, Boca Raton, FL, USA: CRC Press, Inc., 1997.

[38] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," *Bell System Technical Journal*, vol. 61, no. 8, pp. 1871–1894, 1982.

[39] L. Kosten, "Stochastic theory of data handling systems with groups of multiple sources," *Performance of Computer Communication Systems*, pp. 321–331, 1984.

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

[41] H. C. Tijms, *A First Course in Stochastic Models*. West Sussex, England: John Wiley & Sons, Inc., 2003.

[42] V. B. Iversen, "Teletraffic Engineering and Network Planning," 2015.