frontiers | Frontiers in Communications and Networks

TYPE Original Research
PUBLISHED 02 November 2022
DOI 10.3389/frcmn.2022.932111

OPEN ACCESS

EDITED BY
Harpreet S. Dhillon,
Virginia Tech, United States

REVIEWED BY
Praful Mankar,
International Institute of Information Technology, India
Mohamed Abd-Elmagid,
Virginia Tech, United States

*CORRESPONDENCE
Clement Kam,
clement.kam@nrl.navy.mil

SPECIALTY SECTION
This article was submitted to Wireless Communications, a section of the journal Frontiers in Communications and Networks

RECEIVED 29 April 2022
ACCEPTED 23 September 2022
PUBLISHED 02 November 2022

CITATION
Kam C and Kompella S (2022), On the age of information for non-preemptive queues in tandem.
Front. Comms. Net 3:932111.
doi: 10.3389/frcmn.2022.932111

On the age of information for non-preemptive queues in tandem

Clement Kam* and Sastry Kompella

U.S. Naval Research Laboratory, Washington, DC, United States

A monitoring system operates over a network of first-come, first-served queues in tandem, in which a source is transmitting its status to a monitor at the end of the tandem. To characterize the freshness performance of this monitoring system, we analyze the average Age of Information for this system, in which the status update arrivals are Poisson distributed and each queue is served by a non-preemptive, memoryless server. We first study the case of single-capacity queues that are modeled as a stochastic hybrid system, and we derive the average age for two queues with different service rates and three queues with equal service rate. We then study the infinite capacity queue case and use the graphical approach to derive the average age for two queues in tandem with equal service rates. Finally, we simulate the average age for intermediate cases of k capacity queues, which fall in between the two extreme cases of $k = 1$ and $k = \infty$.

KEYWORDS

communications and networking, real-time systems, semantic data and service, sensor and actuator networks, age of information

1 Introduction

In applications for the Internet of Things (IoT) or military applications, it is often important for a device or entity at one point in the network to have information (e.g., sensor data) from another point in the network that is as fresh as possible. The concept of Age of Information (AoI) applies to communication systems where the receiver has an interest in fresh (i.e., most recently generated) information. However, traditional metrics such as packet delay are insufficient for characterizing the performance of systems that rely on real-time monitoring applications. AoI is a widely-explored metric to measure the performance of real-time status updating applications, such as position tracking, environmental/health monitoring, or networked control systems (Kaul et al., 2011; 2012a). In most works studying AoI, the goal is to derive the age averaged over time for various single queue models (Kaul et al., 2012b; Yates and Kaul, 2012; Kam et al., 2013). However, in IoT systems or ad hoc military networks, real-time information is often needed over multiple hops, which can include wireless and wired links. In this work, we focus on analyzing the average AoI for queues in tandem to obtain insight into the performance of real-time status updating systems that operate over a multi-hop network.

Research on AoI has focused on understanding the performance of systems that are modeled by different types of queues, with various arrival/departure processes, number of servers, and queue capacities (Kam et al., 2016a; Bedewy et al., 2016; Costa et al., 2016; Kosta et al., 2017; Sun et al., 2017). While most studies have focused on analyzing single queues, tandem queues have recently been recognized as an important area of study for understanding the AoI for networks beyond a single hop. Chiariotti et al. (2020) studied a system of two connected satellite links for another age-related metric, called the Peak AoI (PAoI), by modeling the system as a series of M/M/1 queues. In contrast to the average AoI studied in this work, PAoI tracks the maximum instantaneous value of AoI for each update. Similarly, tandem satellite links were also studied by Soret et al. (2020), who derived a bound on the average AoI for satellites modeled as queues that received updates directly from ground stations and relayed updates from previous satellites in the tandem. Vikhrova et al. (2020) studied tandem queues with updates that are randomly sent to one of two queues and priority is given to the second queue. In Xu et al. (2020), the average PAoI was analyzed for an IoT network (multiple sources) feeding into tandem queues. In Kuang et al. (2019), a mobile edge computing system was modeled as a tandem queue and the average AoI was analyzed for a zero-wait message generation policy.

There has also been some earlier work on AoI in more general multi-hop networks. For example, wireless scheduling for AoI was studied in Talak et al. (2017), who derived the average and peak average AoI-optimal stationary scheduling policies for line networks, where links are activated according to a stationary probability distribution. However, the authors did not derive the average and peak average AoI in closed form. Bedewy et al. (2017) studied the preemptive Last Generated First Served (LGFS) policy, and show that this policy results in a smaller age at all nodes in the network than any other causal policy. The caveats are that the packets are generated at an external source and are dispersed throughout the network *via* a gateway node, packet transmission times over a network are exponentially distributed, and the age optimization is done in a stochastic ordering sense.

The work that most closely relates to this article is that of Yates (2018); Yates (2020), which studies a line network of preemptive servers. These works analyzed the AoI for this system using a stochastic hybrid system (SHS) modeling approach, as in Yates and Kaul (2019). It was shown in Yates (2020) that for this line network, the age at the destination (or any node in between) is identical in distribution to the sum of independent exponential service times up to the node. Although preemption has been shown to be optimal for exponential servers (Bedewy et al., 2017), the same is not true for general service time distributions. In addition, many practical communication systems do not allow for this type of control for packets in transmission. Furthermore, in most works on queues in tandem, closed form expressions are often given as bounds or for the more tractable PAoI metric. Therefore, it is still important to characterize the average AoI of *non-preemptive* servers in closed form, which is the focus of this article. The analysis here is more challenging than that in the preemption case because it does not allow for the reduction of states in the SHS model, which greatly simplifies the analysis.

More recently, the SHS approach has been extended to not only derive the average AoI but to actually derive the complete distribution of the AoI. The SHS approach for deriving the moment generating function (MGF) for the AoI was first presented in Yates (2020), where it was also applied to a line network of preemptive queues. However, few works have since applied that approach to different systems, such as an energy-harvesting node with varying types of preemptive and non-preemptive policies by Abd-Elmagid and Dhillon (2022a); Abd-Elmagid and Dhillon (2022b), and a two-source single-server system with packet management by Moltafet et al. (2021). In this work, we have derived the first moment (average) of the AoI for non-preemptive queues in tandem. However, deriving the MGF is a more complex derivation that is being considered for a follow-up work.

Koukoutsidis (2020) developed an approach to calculate the average AoI for an overtake-free network of quasi-reversible queues, including M/M/1/∞ queues in tandem (which we also consider), as well as networks with different classes of update packets. In contrast to the SHS approach, they used the queueing theoretical result that the end-to-end sojourn time for an overtake-free path is distributed as a sum of independent exponential sojourns at each node. However, in their AoI calculation, they make the assumption that, because the interarrival and interdeparture times are identically distributed, their correlation with the sojourn time is equal[1]. This is not true in general and thus they derive a different result for the M/M/1/∞ queues in tandem[2].

In this work, we start by deriving the average age for two non-preemptive queues in tandem, in which each of the queues have a capacity equal to 1. This is done using the SHS approach. Limiting the queue capacity is based on the intuition that allowing updates to age in a queue may not be efficient when the objective is to maximize information freshness at the destination because the queues will not store obsolete packets unnecessarily. This has the additional advantage of making the analysis of the average age more tractable. Furthermore, this should approximate the performance of a non-preemptive LGFS

1 In Koukoutsidis (2020), they incorrectly assume $E[S_i^{c,1\to n}D_i^c] = E[S_i^{c,1\to n}A_i^c]$, where for a packet class c, $S_i^{c,1\to n}$ is the sojourn time for the ith packet, D_i^c is the interdeparture time between packet i and i + 1, and A_i^c is the interarrival time between packet i − 1 and i.

2 We have verified that simulation results do not match the numerical results from the expression in Koukoutsidis (2020) but the simulations validate our expression.

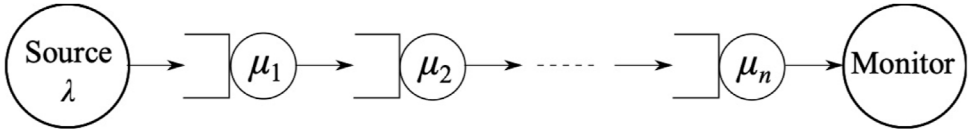

FIGURE 1
Queues in tandem.

and is likely to outperform it. To verify our theoretical result, we simulate the single-capacity non-preemptive queues in tandem. We extend the analysis by considering three non-preemptive queues in tandem. We then compare our result with an approximation for arbitrary number of non-preemptive single-capacity queues in tandem.

Next, we consider the case of infinite capacity queues in tandem, with a Poisson arrival process and exponential servers. Unfortunately, applying the SHS approach in this case makes the problem intractable because this results in an explosion of states in the SHS model. Therefore, we resort to the traditional method of deriving the average age using a graphical argument, whereby we evaluate the correlations between system times and interarrival or interdeparture times. Again, to verify the theoretical result, we simulate the infinite capacity non-preemptive queues in tandem. Finally, to develop a holistic understanding of this problem space, we study the case of capacity-k non-preemptive queues in tandem. We generate simulation results for capacity-k queues in tandem because we lack theoretical expressions of average age for arbitrary values of k.

The remainder of this article is organized as follows. The system model is described in Section 2. The average age for capacity-1 queues in tandem is derived in Section 3 using the SHS approach. The average age for infinite capacity queues in tandem is derived in Section 4 using the traditional graphical approach. Various simulation results, including those for capacity-k queues, are provided in Section 5. Finally, we summarize our results in Section 6.

2 System model

We study the AoI over a network of queues in tandem, as shown in Figure 1. A source generates arrivals as a rate λ Poisson process and then sends them to the first queue. The packets flow through a series of queues serving packets at rate μ_i until they reach the monitor (as depicted in the figure). There is no preemption in this system, such that a packet in service completes service before the server becomes available for another packet and a packet that arrives to a full queue is dropped. Aside from this queue overflow, there are no other packet losses under this model.

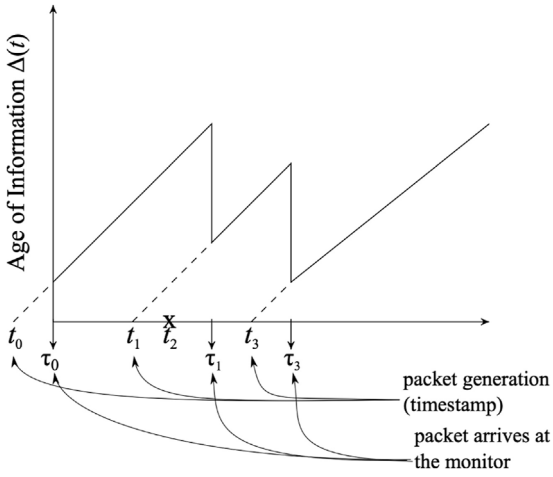

FIGURE 2
Age of information vs. time, single-capacity queue: t_i and τ_i are the time of generation and time of arrival at the monitor of the ith update, respectively.

The AoI at the monitor at time t is defined as $\Delta(t) = t - u(t)$, where $u(t)$ is the time at which the freshest packet at the monitor was generated at the source. A sample path of the age for a single-capacity queue is given in Figure 2, where t_i and τ_i are the time of generation and time of arrival at the monitor of the ith update, respectively. In this figure, packet 2 is dropped due to the server being occupied. We are interested in deriving the average AoI,

$$\Delta = \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} \Delta(t)\mathrm{d}t. \tag{1}$$

3 Average age of single-capacity queues in tandem

3.1 Two ·/M/1/1 queues in tandem

For our theoretical analysis, we start by looking at two queues with a capacity of one, such that any packet that arrives to a queue with a packet already in service is dropped. We focus on a

Poisson arrival process from the source into the first queue and the packets are served at each queue with an exponential service time. The traditional method of deriving the average AoI involves a graphical argument and the need to evaluate correlations between system times and interarrival or interdeparture times (Kaul et al. 2012a). With queues in tandem, this approach is non-trivial—not only in calculating the correlations between the waiting times and the interarrival times but it also requires the correlation between waiting times at the different queues to be calculated, among other correlations. In a capacity-limited system, we would also need to account for packet losses at both queues. In this section, we use the notation ·/M/1/1, where the initial "·" indicates that there are no exogenous arrivals to the queues further down the tandem.

An alternative method of calculating the average age was recently proposed that models the system as a SHS (Yates and Kaul, 2019), which avoids the complexity of computing the correlation quantities in the graphical approach. The SHS is modeled with a hybrid state $(q(t), \mathbf{x}(t))$, where $q(t) \in \mathcal{Q}$ is the discrete Markov state of the queueing system and $\mathbf{x}(t) \in \mathbb{R}^n$ is the continuous state that captures the evolution of the ages at different points in the system. Following the approach in Yates (2018), we model the two ·/M/1/1 queues in tandem (as described in Section 2), with the discrete state set $\mathcal{Q} = \{00, 10, 01, 11\}$, where $q_1q_2 \in \mathcal{Q}$, $q_i = 1$ indicates that queue i has a packet in service, and $q_i = 0$ when the queue is empty[3]. The continuous state for this system is $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]$, where $x_3(t)$ is the age at the monitor and $x_i(t)$ is the age of the packet being served at queue i when there is a packet in service, and otherwise $x_i(t)$ is irrelevant and set to 0.

For the Markov chain $q(t)$, the transitions between states $l \in \mathcal{L}$ are directed edges (q_l, q_l') with transition rate $\lambda^l \delta_{q_l, q(t)'}$, where the Kronecker delta ensures that the transition l only occurs in state q_l. In SHS, these transitions correspond to changes in the discrete state $q(t)$, as well as jumps in the continuous state $\mathbf{x}(t)$, according to $\mathbf{x}' = \mathbf{x}\mathbf{A}_l$, where $\mathbf{A}_l$ is a binary matrix that defines the jumps in $\mathbf{x}(t)$ for the discrete state transition l. These transitions may be self-transitions, in which the discrete state stays the same but the continuous state jumps. In addition, there may be multiple transitions between the same pair of states depending on the different impact on the continuous state, which is unlike typical continuous-time Markov chains. In this work, there are no self-transitions or multiple transitions that need to be modeled.

For all $\hat{q} \in \mathcal{Q}$, we define $\pi_{\hat{q}}(t)$ to be the discrete Markov state probabilities. We define $v_{\hat{q}j}(t)$ as the correlation between the continuous age state $\mathbf{x}(t)$ and the occupancy of the discrete state at time t; that is, $q(t)$. These definitions are given as follows:

$$\pi_{\hat{q}}(t) = \mathbb{E}\left[\delta_{\hat{q},q(t)}\right] \tag{2}$$

$$v_{\hat{q}j}(t) = \mathbb{E}\left[x_j(t)\delta_{\hat{q},q(t)}\right], \quad j \in \{1, \ldots, n\}. \tag{3}$$

From this, we also have the vector functions

$$\mathbf{v}_{\hat{q}}(t) = \left[v_{\hat{q}1}(t), \ldots, v_{\hat{q}n}(t)\right] = \mathbb{E}\left[\mathbf{x}(t)\delta_{\hat{q},q(t)}\right], \tag{4}$$

where n is the number of dimensions in the continuous state $\mathbf{x}(t)$. For an ergodic queueing system, the average age is given by $\Delta = \mathbb{E}[x_n] = \sum_{\bar{q}\in\mathcal{Q}} \bar{v}_{\bar{q}n}$, where $\bar{v}_{\bar{q}j} = \lim_{t\to\infty} \bar{v}_{\bar{q}j}(t)$ is the steady-state correlation between the discrete and age states. To derive the steady state $\mathbf{v}_{\bar{q}}(t)$, we need the following first order differential equation for all $\bar{q} \in \mathcal{Q}$ as derived in Yates and Kaul (2019):

$$\dot{\mathbf{v}}_{\bar{q}}(t) = \mathbf{1}\bar{\pi}_{\bar{q}} + \sum_{l\in\mathcal{L}'_{\bar{q}}} \lambda^{(l)} \mathbf{v}_{q_l}(t)\mathbf{A}_l - \mathbf{v}_{\bar{q}}(t) \sum_{l\in\mathcal{L}_{\bar{q}}} \lambda^{(l)} \tag{5}$$

where $\mathcal{L}'_{\bar{q}}$ is the set of transitions entering $\bar{q}$ and $\mathcal{L}_{\bar{q}}$ is the set of transitions leaving $\bar{q}$. The rate at which the continuous state changes is the $\mathbf{1}$ vector because the age increases linearly with a slope of 1. To solve for the steady state $\mathbf{v}_{\bar{q}}(t)$, we let $\dot{\mathbf{v}}_{\bar{q}}(t) = 0$, yielding

$$\bar{\mathbf{v}}_{\bar{q}} \sum_{l\in\mathcal{L}_{\bar{q}}} \lambda^{(l)} = \mathbf{1}\bar{\pi}_{\bar{q}} + \sum_{l\in\mathcal{L}'_{\bar{q}}} \lambda^{(l)} \bar{\mathbf{v}}_{q_l}\mathbf{A}_l, \quad \bar{q} \in \mathcal{Q}. \tag{6}$$

According to Yates and Kaul (2019, Theorem 4), if we can find a non-negative solution $\bar{\mathbf{v}} = [\bar{\mathbf{v}}_{00} \cdots \bar{\mathbf{v}}_{11}]$ to (6), then the then differential Eq. 5 is stable and the average age is given by $\Delta = \sum_{\bar{q}\in\mathcal{Q}} \bar{v}_{\bar{q}3}$.

For the two single-capacity queue system in this section, the discrete Markov chain is given in Figure 3, where the states are $q \in \mathcal{Q}$ and the transition labels are $l \in \mathcal{L}$, which correspond to $q_l \to q_l'$. Following Yates and Kaul (2019), we list the transitions,

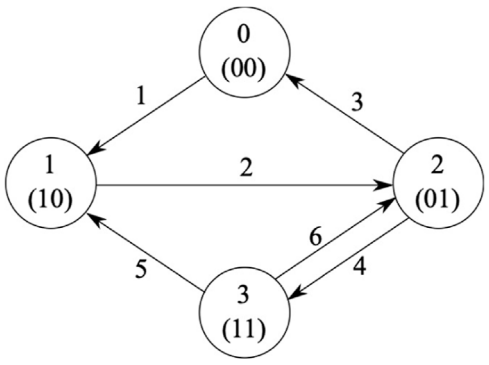

FIGURE 3
SHS Markov chain for two ·/M/1/1 queues in tandem.

3 In Yates (2018), the authors were able to simplify the derivation for any number of preemptive queues in tandem by using a "fake updates" trick to reduce the number of states to a single state. However, this trick does not work in the system here with non-preemptive queues. Consequently, we have the number of states being 2^m, where m is the number of queues.

TABLE 1 Table of transitions for the Markov chain in Figure 3.

l	$q_l \to q_l'$	$\lambda^{(l)}$	$\mathbf{xA}_l$	$\mathbf{A}_l$	$\overline{\mathbf{v}_{ql}}\mathbf{A}_l$
1	$00 \to 10$	λ	$[0 \quad 0 \quad x_3]$	$\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$	$[0 \quad 0 \quad \bar{v}_{003}]$
2	$10 \to 01$	μ_1	$[0 \quad x_1 \quad x_3]$	$\begin{bmatrix}0&1&0\\0&0&0\\0&0&1\end{bmatrix}$	$[0 \quad \bar{v}_{101} \quad \bar{v}_{103}]$
3	$01 \to 00$	μ_2	$[0 \quad 0 \quad x_2]$	$\begin{bmatrix}0&0&0\\0&0&1\\0&0&0\end{bmatrix}$	$[0 \quad 0 \quad \bar{v}_{012}]$
4	$01 \to 11$	λ	$[0 \quad x_2 \quad x_3]$	$\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}$	$[0 \quad \bar{v}_{012} \quad \bar{v}_{013}]$
5	$11 \to 10$	μ_2	$[x_1 \quad x_2 \quad x_2]$	$\begin{bmatrix}1&0&0\\0&1&1\\0&0&0\end{bmatrix}$	$[\bar{v}_{111} \quad \bar{v}_{112} \quad \bar{v}_{112}]$
6	$11 \to 01$	μ_1	$[0 \quad x_2 \quad x_3]$	$\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}$	$[0 \quad \bar{v}_{112} \quad \bar{v}_{113}]$

the transition rates, and the quantities in (6) associated with the mapping of jumps in the continuous state at transitions $\mathbf{A}_l$ in Table 1.

Two examples of how we determined the table entries follow. Transition $l = 1$ corresponds to a packet arrival to a completely empty system, which occurs at rate λ, and the jump in the continuous state ($\mathbf{xA}_l$) is such that the age at the monitor remains the same (x_3), the packet in service at queue 1 starts at 0, and the age at the empty server does not matter, so it is set to 0. A transition that is specific to non-preemptive queues in tandem is $l = 6$, which corresponds to the packet in the first server completing service (at a rate of μ_1) in a full system, such that it gets dropped at the second queue. For $\mathbf{xA}_l$, the age at the monitor remains the same (x_3), the age at the second queue remains the same because there is no preemption (x_2), and the age at the first monitor goes to zero because it was served and dropped at the second queue.

We use the balance equations to derive the discrete Markov state stationary probabilities $\bar{\pi}_{\bar{q}}$:

$$\bar{\pi}_{\bar{q}} \sum_{l \in \mathcal{L}_{\bar{q}}} \lambda^{(l)} = \sum_{l \in \mathcal{L}'_{\bar{q}}} \lambda^{(l)} \bar{\pi}_{q_l}, \quad \bar{q} \in \mathcal{Q}$$
$$\sum_{\bar{q} \in \mathcal{Q}} \bar{\pi}_{\bar{q}} = 1.$$

Filling these equations with the quantities from Table 1 and by solving for the stationary probabilities, we obtain.

$$\bar{\pi}_{00} = \frac{\mu_1 \mu_2}{(\lambda + \mu_1)(\lambda + \mu_2)} \tag{7a}$$

$$\bar{\pi}_{01} = \frac{\lambda \mu_1}{(\lambda + \mu_1)(\lambda + \mu_2)} \tag{7b}$$

$$\bar{\pi}_{10} = \frac{\lambda \mu_2 (\lambda + \mu_1 + \mu_2)}{(\lambda + \mu_1)(\lambda + \mu_2)(\mu_1 + \mu_2)} \tag{7c}$$

$$\bar{\pi}_{11} = \frac{\lambda^2 \mu_1}{(\lambda + \mu_1)(\lambda + \mu_2)(\mu_1 + \mu_2)}. \tag{7d}$$

We can substitute these stationary probabilities and the quantities in Table 1 into (6) to obtain the following system of equations:

$$\lambda \bar{v}_{003} = \pi_{00} + \mu_2 \bar{v}_{012} \tag{8a}$$
$$(\lambda + \mu_2)\bar{v}_{013} = \pi_{01} + \mu_1 (\bar{v}_{103} + \bar{v}_{113}) \tag{8b}$$
$$\mu_1 \bar{v}_{103} = \pi_{10} + \lambda \bar{v}_{003} + \mu_2 \bar{v}_{112} \tag{8c}$$
$$(\mu_1 + \mu_2)\bar{v}_{113} = \pi_{11} + \lambda \bar{v}_{013} \tag{8d}$$
$$(\lambda + \mu_2)\bar{v}_{012} = \pi_{01} + \mu_1 (\bar{v}_{101} + \bar{v}_{112}) \tag{8e}$$
$$(\mu_1 + \mu_2)\bar{v}_{112} = \pi_{11} + \lambda \bar{v}_{012} \tag{8f}$$
$$\mu_1 \bar{v}_{101} = \pi_{10} + \mu_2 \bar{v}_{111} \tag{8g}$$
$$(\mu_1 + \mu_2)\bar{v}_{111} = \pi_{01} \tag{8h}$$

To derive the average age, it is not necessary to solve for all of the $\bar{v}_{\bar{q}j}$ but only $\sum_{\bar{q} \in \mathcal{Q}} \bar{v}_{\bar{q}3}$. We start by finding the linear combination of Eqs 8a–8d using coefficients a, b, c, and d, respectively, which yields

$$a\lambda \bar{v}_{003} + b(\lambda + \mu_2)\bar{v}_{013} + c\mu_1 \bar{v}_{103} + d(\mu_1 + \mu_2)\bar{v}_{113}$$
$$= a(\pi_{00} + \mu_2 \bar{v}_{012}) + b(\pi_{01} + \mu_1 (\bar{v}_{103} + \bar{v}_{113}))$$
$$+ c(\pi_{10} + \lambda \bar{v}_{003} + \mu_2 \bar{v}_{112}) + d(\pi_{11} + \lambda \bar{v}_{013})$$

We want to set the coefficients such that moving all $\bar{v}_{\bar{q}3}$ terms to the left-hand side yields $\sum_{\bar{q} \in \mathcal{Q}} \bar{v}_{\bar{q}3}$; that is, find a, b, c, and d that satisfies

$$\lambda a - \lambda c = 1$$
$$(\lambda + \mu_2)b - \lambda d = 1$$
$$\mu_1 c - \mu_1 b = 1$$
$$(\mu_1 + \mu_2)d - \mu_1 b = 1$$

which has solution

$$a = \frac{1}{\lambda} + \frac{1}{\mu_1} + \frac{1}{\mu_2}, \qquad b = \frac{1}{\mu_2}, \qquad c = \frac{1}{\mu_1} + \frac{1}{\mu_2}, \qquad d = \frac{1}{\mu_2}.$$

Combining Eqs 8a–8d with respective coefficients a, b, c, and d yields

$$\sum_{\bar{q} \in \mathcal{Q}} \bar{v}_{\bar{q}3} = a\pi_{00} + b\pi_{01} + c\pi_{10} + d\pi_{11} + a\mu_2 \bar{v}_{012} + c\mu_2 \bar{v}_{112}.$$

To find $a\mu_2 \bar{v}_{012} + c\mu_2 \bar{v}_{112}$, we again take linear combinations of Eqs 8e,8f, and we need to find the coefficients a' and b' that satisfy

$$(\lambda + \mu_2)a' - \lambda b' = a\mu_2 (\mu_1 + \mu_2)b' - \mu_1 a' = c\mu_2$$

which has solution

$$a' = \frac{\mu_1 + \mu_2}{\lambda(\lambda + \mu_1 + \mu_2)} + \frac{1}{\mu_1} + \frac{1}{\mu_2}, \qquad b' = \frac{1}{\mu_1} + \frac{\mu_1}{\mu_1 + \mu_2} a'.$$

The last linear combination of equations is for (Eq. 8g) and (Eq. 8h) to find $\mu_1 a' \bar{\nu}_{101}$, and the coefficients satisfy

$$\mu_1 a'' = \mu_1 a' (\mu_1 + \mu_2) b'' - \mu_2 a'' = 0$$

with solution

$$a'' = a', \qquad b'' = \frac{\mu_2}{\mu_1 + \mu_2} a'.$$

Finally, we have

$$\Delta = a\pi_{00} + (b + a')\pi_{01} + (c + a'')\pi_{10} + (d + b' + b'')\pi_{11}$$

After much simplification, we have our first result:

Theorem 1. *The average age after two single-capacity queues in tandem (M/M/1/1 → ·/M/1/1) is given by*

$$\Delta = \frac{1}{\lambda} + \frac{1}{\mu_1} + \frac{1}{\mu_2} + \frac{\lambda}{\mu_1(\lambda + \mu_1)} + \frac{\lambda}{\mu_2(\lambda + \mu_2)}.$$

We can see from the result that the average age is insensitive to the ordering of the servers. Next, we consider the case where the service rates at both queues are the same ($\mu_1 = \mu_2 = \mu$). The average age in this case is given by

$$\Delta_{M/M/1/1,2} = \frac{1}{\lambda} + \frac{2}{\mu}\left(1 + \frac{\lambda}{\lambda + \mu}\right). \tag{9}$$

Noting the similarity to the average age for a single M/M/1/1 Costa et al. (2016), we have for m queues in tandem,

$$\Delta_{M/M/1/1,m} = \frac{1}{\lambda} + \frac{m}{\mu}\left(1 + \frac{\lambda}{\lambda + \mu}\right), \tag{10}$$

for $m = 1, 2$. We will consider this as an approximation for $m > 2$. We will show that this expression does not hold for $m = 3$. As $\lambda \to \infty$, the average age approaches $2m/\mu$ for $m = 1, 2$.

If we take the derivative of (Eq. 9) with respect to λ, then we have

$$\frac{\partial \Delta}{\partial \lambda} = \frac{\lambda^2 - 2\lambda\mu - \mu^2}{\lambda^2 (\lambda + \mu)^2}.$$

The minimum occurs at

$$\lambda_{\min} = \mu(1 + \sqrt{2}) \approx 2.414\mu \tag{11}$$

It was shown in Kam et al. (2016b) that the average age for a single M/M/1/1 queue was decreasing for all λ, whereas the average age for a single M/M/1/2 reached a minimum at $\lambda \approx 1.427$ (for $\mu = 1$). In this case, adding another queue in tandem of capacity 1 after the M/M/1/1 queue also has the behavior of meeting a minimum but at a higher λ. The value of the minimum age after two queues is

$$\Delta_{M/M/1/1,2,\min} = \frac{4}{\mu} - \frac{(\sqrt{2} - 1)^2}{\mu}. \tag{12}$$

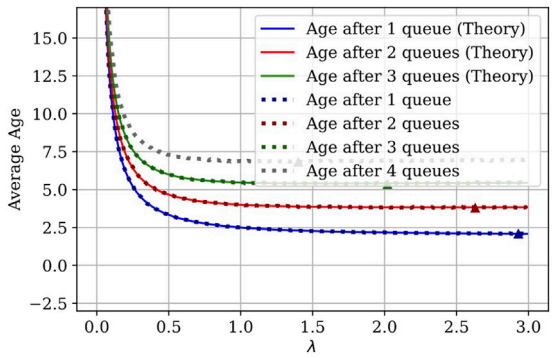

FIGURE 4
Average age vs. λ, μ = 1, queue capacity of one, without preemption. Simulation results are indicated by the points, theoretical results are indicated by the solid line, and simulation minima are indicated by △.

If we compare the minimum age to the age as $\lambda \to \infty$, then the penalty for using an infinite λ is $(\sqrt{2} - 1)^2/\mu$, which is about $0.17/\mu$.

Using the expression for the average age for the M/M/1/2 in Costa et al. (2016), we can show that the average age for the queues in tandem in (11) is greater than that of the M/M/1/2. This is intuitive due to having to go through two servers instead of just one.

3.2 Three (equal rate) ·/M/1/1 queues in tandem

In this section, we build on the previous case and now consider three queues in tandem, for the case where the servers have equal service rate. Again we use the stochastic hybrid systems approach and the discrete state set is $\mathcal{Q} = \{000, 001, 010, 011, 100, 101, 110, 111\}$, where $q_1 q_2 q_3 \in \mathcal{Q}$, $q_i = 1$ indicates that queue i has a packet in service, and $q_i = 0$ indicates when the queue is empty. Using the same approach for the two-queue case, we derive the following expression for the average age:

Theorem 2. The average age after three single-capacity queues in tandem (M/M/1/1 → ·/M/1/1 → ·/M/1/1) is given by

$$\Delta_{M/M/1/1,3} = \frac{1}{\lambda} + \frac{3}{\mu} + \frac{8\lambda^3 + 18\lambda^2\mu + 24\lambda\mu^2}{9\mu(\lambda + \mu)^3}$$
$$= \frac{1}{\lambda} + \frac{3}{\mu} + \frac{3\lambda}{\mu(\lambda + \mu)} - \frac{\lambda^3}{3\mu(\lambda + \mu)^3}$$

Proof. The proof is given in the Appendix.

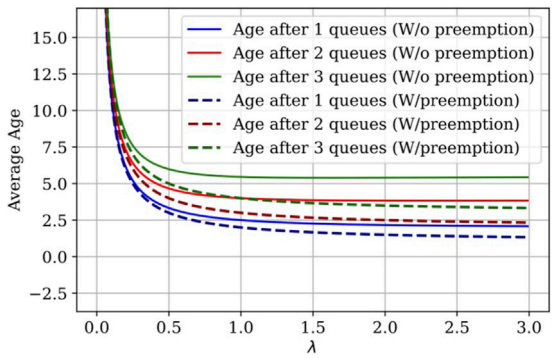

FIGURE 5
Average age vs. λ, μ = 1, queue capacity of one, preemption vs. non-preemption.

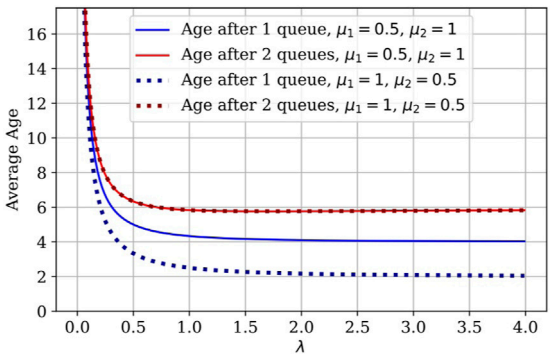

FIGURE 7
Average age vs. λ, different μ_1, μ_2 queue capacity of one, non-preemption.

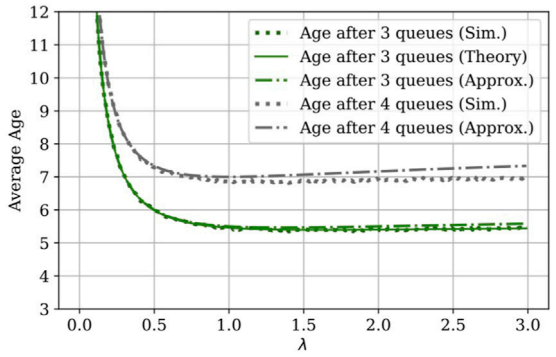

FIGURE 6
Average age vs. λ, μ = 1, queue capacity of one after three and four queues (vs. approximation).

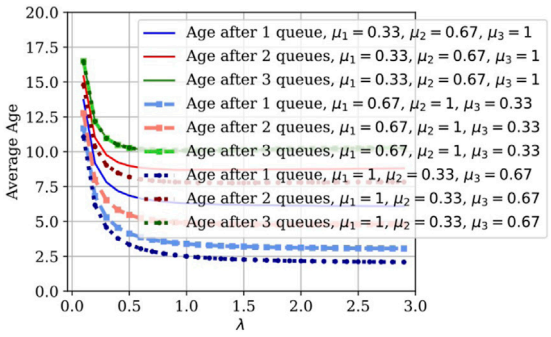

FIGURE 8
Average age vs. λ, different μ_1, μ_2, μ_3, queue capacity of one, non-preemption.

If we apply the expression in (Eq. 10) to m = 3, then we see that this can serve as an approximation for the average age that is greater by $\frac{\lambda^3}{3\mu(\lambda+\mu)^3}$.

3.3 Numerical/simulation results

To verify our theoretical results, we look at the ·/M/1/1 queues with Poisson arrivals for up to four queues in tandem. The results for the average age as a function of λ are given in Figure 4. For comparison, we plot the numerically evaluated theoretical expressions for the average age up to three queues in tandem, and we observe that the plot lines for the theoretical and simulated results lie on top of one another. We also plot the minima from simulation and compare with (Eq. 11), which evaluates to 2.414 for $m = 2$. This matches with the simulation. We observe that the minimum λ is decreasing as the number of queues increases. For comparison, the case with preemption in Yates (2022) is shown alongside the non-preemptive case in Figure 5. The age for the preemptive case is lower than that of the non-preemptive case by about 40% at each point in the tandem network. We also consider the expression in (Eq. 10), which is exact for m = 1, 2 but is an approximation for m = 3, 4. Figure 6 shows the simulated, theoretical, and approximate results for m = 3, 4. The approximation overestimates the average age with a gap that increases with λ. For large λ, the approximation is only off by about 5% in the three-queue case. However, it is closer to 7% in the four-queue case, which suggests that the approximation may not scale well.

We also numerically evaluate (13) for a two-queue system with different service rates, 0.5 and 1, and the average age is plotted in Figure 7. It can be seen that the plot lines for the average age after two queues are right on top of each other, which demonstrates that the average age after the second queue is indeed insensitive to the

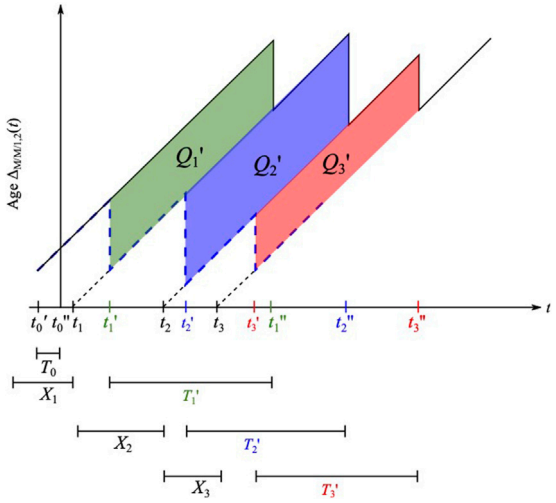

FIGURE 9
Age computation at (M/M/1) queue n based on age at queue n − 1.

ordering of the servers (as previously stated). To further validate this claim, we simulate a three-queue system with different service rates; that is, 0.33, 0.67, and 1. We plot the average age for different permutations of service rates in Figure 8. The average age after three queues is the same for all three permutations, which is further proof that the order of servers in the tandem does not impact the average age after the last queue.

4 Average age of two ·/M/1/∞ queues in tandem

We next consider the case of infinite capacity first-come, first-served queues in tandem with Poisson arrival process and exponential servers. To derive the age after two queues, we use a graphical approach as in Kaul et al. (2012a) to derive the average age going after the second queue based on the average age after the first queue. We will show that for each packet i, the additional area under the age function beyond the age after the first queue is XT', where X is the interarrival time at the first queue and T' is the system time in the second queue.

Theorem 3. The average age after two infinity capacity queues in tandem (M/M/1/∞ → ·/M/1/∞) is given by

$$\Delta_{M/M/1,2} = \Delta_{M/M/1,1} + \lambda\left(E\left[XT'\right]\right) = \frac{1}{\lambda} + \frac{1}{\mu_1} + \frac{\lambda^2}{\mu_1^2(\mu_1-\lambda)} + \frac{1}{\mu_2} + \frac{\lambda^2}{2\mu_2-\lambda} + \frac{\lambda^2}{\mu_1\mu_2(\mu_1+\mu_2-\lambda)}. \quad (13)$$

Proof. An example of the evolution of the age function after queue 2 is shown in Figure 9. To derive the average age (1) using the graphical method, we calculate the total area under the age function by breaking it up into geometric shapes and then calculate their areas separately. The age function after queue 1, $\Delta_{M/M/1,1}(t)$, is shown in the figure by a dotted line, which is bounded above by the age after queue 2, $\Delta_{M/M/1,2}(t)$. We seek to calculate the remaining area under the age function after accounting for the average age $\Delta_{M/M/1,1}$.

At time t_0', the packet is served at queue 1 and arrives at queue 2. At the origin of the plot (t_0''), the packet is served at queue 2 and the age after queue 2 ($\Delta_{M/M/1,2}$) is defined. From t_0'' until another packet arrives and is served at queue 1 (t_1'), the ages after both queues are the same. At the service time t_1' at queue 1, the age $\Delta_{M/M/1,1}(t)$ resets to the lower age but the age $\Delta_{M/M/1,2}(t)$ does not yet reset to the age of the packet until it is served at t_1''. The amount that the age is reduced by is equal to the difference between the ages of consecutive packets, which is equal to the interarrival time X_1 into queue 1. The area of the parallelogram is given by the X_1 times the system time in queue 2, $T_1' = t_1'' - t_1'$.

The average age at queue 2 is given by

$$\begin{aligned}\Delta_{M/M/1,2} &= \lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}}\Delta_{M/M/1,2}(t)\mathrm{d}t \\ &= \lim_{\mathcal{T}\to\infty}\left[\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}}\Delta_{M/M/1,1}(t)\mathrm{d}t + \frac{R(\mathcal{T})}{\mathcal{T}} + \frac{N(\mathcal{T})}{\mathcal{T}}\sum_{i=1}^{N(\mathcal{T})}\frac{T_i'X_i}{N(\mathcal{T})}\right] \\ &= \Delta_{M/M/1,1} + \lambda E\left[T'X\right]\end{aligned} \quad (14)$$

where $N(\mathcal{T})$ is the number of update packets served before time $\mathcal{T}$ and $R(\mathcal{T})$ is the residual area under the curve less the area under $\Delta_{M/M/1,1}(t)$. The arrival rate into queue 2 is λ and $R(\mathcal{T})/\mathcal{T}$ disappears in the limit, provided $\lambda < \mu_2$. The age after the first queue $\Delta_{M/M/1,1}$ can be found in Kaul et al. (2012a).

To derive E[T′X], we have $E[T_i'X_i] = E[X_i(W_i' + S_i')] = E[W_i'X_i] + E[X]E[S]$. The waiting time can be written as $W_i' = (T_{i-1}' - Y_i)^+$, where $(x)^+ = max(0, x)$ and Y_i is the interdeparture time of the first queue. We consider two cases for the interdeparture time. If the interarrival time for packet i is less than the system time for packet i − 1 ($X_i \le T_{i-1}$), then packet i immediately begins service and the interdeparture time is $Y_i = S_i$. Otherwise, $Y_i = X_i - T_{i-1} + S_i$. By combining these two cases, we can write $Y_i = (X_i - T_{i-1})^+ + S_i$. The conditional expected waiting time at the second queue W_i' given the interarrival time at the first queue $X_i = x$ is given by

$$\begin{aligned}E\left[W_i'|X_i = x\right] &= E\left[\left(T_{i-1}' - Y_i\right)^+|X_i = x\right] \\ &= E\left[\left(T_{i-1}' - (x - T_{i-1})^+ - S_i\right)^+\right].\end{aligned}$$

Note that T_{i-1}' and T_{i-1} are independent of the service time and interarrival time of future packets, including S_i and X_i. In addition, for the pair of M/M/1/∞ → ·/M/1/∞ queues in tandem, the system times T_i and T_i' are independent of one another (Karpelevitch and Kreinin (1992)). By Burke's Theorem (Burke (1956)), the departure process from the first queue is Poisson with rate λ, so the second queue can also be viewed as an

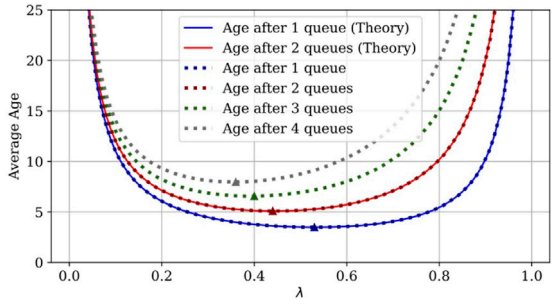

FIGURE 10
Average age vs. λ, μ = 1, infinite queue capacity.

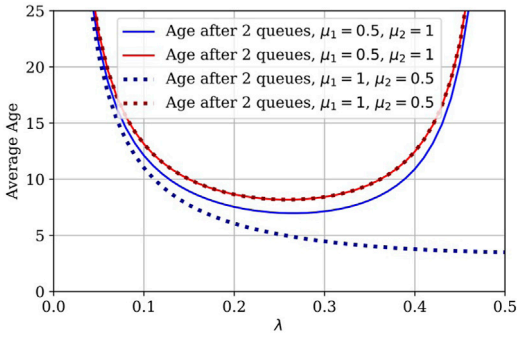

FIGURE 11
Average age vs. λ, different μ_1, μ_2, infinite queue capacity.

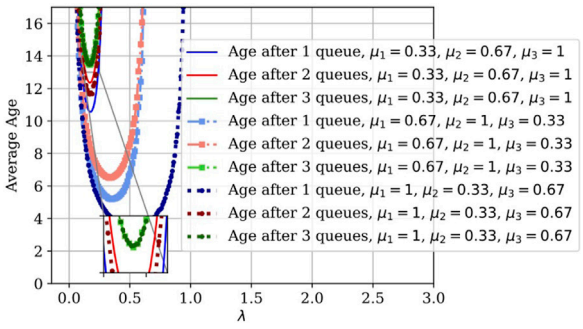

FIGURE 12
Average age vs. λ, different μ_1, μ_2, μ_3, infinite queue capacity.

independent M/M/1 queue and the service time distribution for each queue is given by $f_T(t) = (\mu_2 - \lambda)e^{-(\mu_2-\lambda)t}$. Therefore, we have

$$E[W_i'|X_i = x] = \int_0^\infty \int_0^x \int_{x+s-t}^\infty (t' - x + t - s) f_T(t') f_T(t) f_S(s) \mathrm{d}t' \mathrm{d}t \mathrm{d}s + \int_0^\infty \int_x^\infty \int_s^\infty (t' - s) f_T(t') f_T(t) f_S(s) \mathrm{d}t' \mathrm{d}t \mathrm{d}s = \frac{\mu_1}{(\mu_2 - \lambda)(\mu_1 - \mu_2)(\mu_1 + \mu_2 - \lambda)} \left((\mu_1 - \lambda)e^{-(\mu_2-\lambda)x} - (\mu_2 - \lambda)e^{-(\mu_1-\lambda)x}\right). \quad (15)$$

Using (Eq. 15), we can solve for E[W′X]:

$$E[W'X] = \int_0^\infty xE[W_i'|X_i = x] f_X(x) \mathrm{d}x = \int_0^\infty \left[\frac{\mu_1}{(\mu_2 - \lambda)(\mu_1 - \mu_2)(\mu_1 + \mu_2 - \lambda)} \left((\mu_1 - \lambda)xe^{-\mu_2 x} - (\mu_2 - \lambda)xe^{-\mu_1 x}\right)\right] \mathrm{d}x = \frac{\lambda}{\mu_2^2(\mu_2 - \lambda)} + \frac{\lambda}{\mu_1\mu_2(\mu_1 + \mu_2 - \lambda)}.$$

Finally, we can substitute this into λE[T′X] = λ(E[W′X] + EXES) = λE[W′X] + 1/μ_2, which can be substituted into (Eq. 14) to complete the proof.

From Eq. 14, we can see that even when the service rates are different, the average age does not depend on the order of the servers. Also recall that the order of servers did not affect the average age for single-capacity queues.

This proof for two infinite capacity queues can be extended beyond two queues. However, this would be cumbersome because of the calculation of the conditional waiting time at the last queue given the interarrival time at the first queue and would require the cases where the interarrival/interdeparture times are longer or shorter than the system times to be tracked.

We evaluated the theoretical result for 1 M/M/1/∞ and the M/M/1/∞ → ·/M/1/∞ (Eq. 13), and then simulated for up to four queues in tandem; the results are shown in Figure 10. The plot lines for the simulations and numerically evaluated expressions after 1 and 2 queues lie on top of one another. We also include the minimum for each curve on the plot (indicated by △) and we see that the λ that minimizes the average age again decreases as the number of queues increases. Under these settings with Poisson arrival and exponential service times, the performance can be improved by backing off from the previously optimal packet generation rate λ as the number of queues increases. In Figure 11, we plot the age for two cases of heterogeneous queues in tandem: a) where $\mu_1 = 0.5$ and $\mu_2 = 1$ (solid line), and b) where $\mu_1 = 1$ and $\mu_2 = 0.5$ (dotted line). As we noted from the expression in Eq. 14, the order of the servers does not affect the average age after the second queue. We simulated this for three cases of three queues in Figure 12 and we observe that the average age after the third queue is the same in all three cases (and for all other permutations of queue orders not shown here).

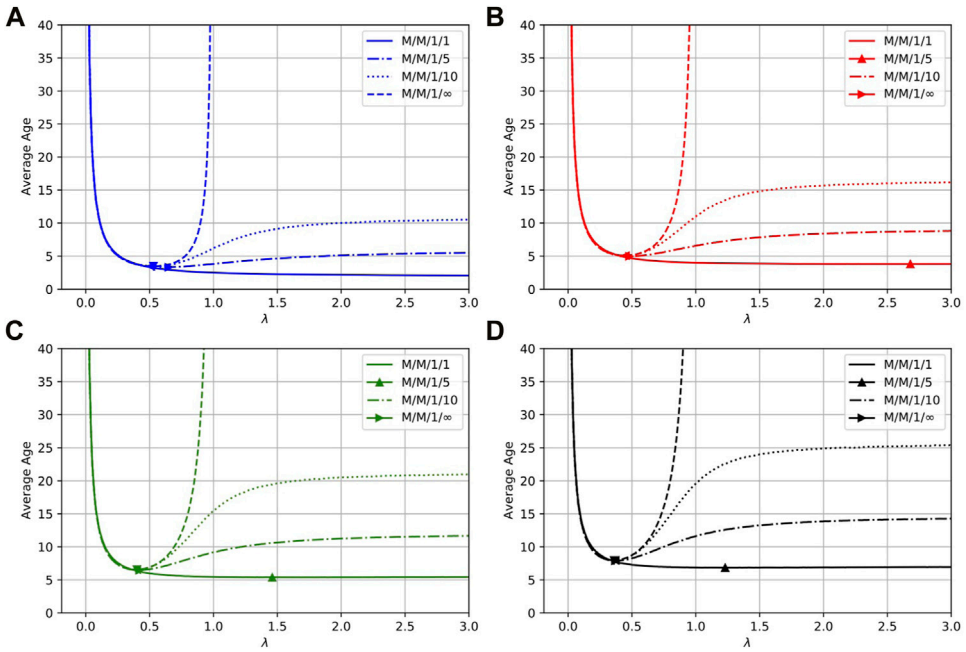

FIGURE 13
Average age vs. \λ, \μ = 1, k queue capacity. **(A)** 1 queue, **(B)** 2 queues in tandem, **(C)** 3 queues in tandem, **(D)** 4 queues in tandem.

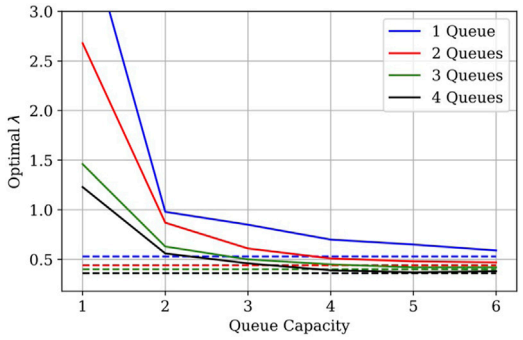

FIGURE 14
Optimal λ vs. queue capacity, μ = 1. The dotted line is the optimal λ for infinite capacity queues.

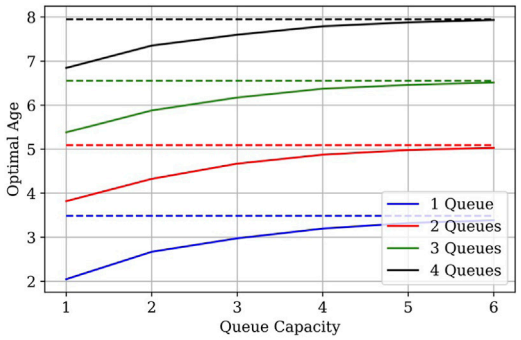

FIGURE 15
Optimal average age vs. queue capacity, μ = 1. The dotted line is the optimal age for infinite capacity queues.

5 Simulation results for k capacity queues

To study the regime between single-capacity and infinite capacity queues in tandem, we simulate the average AoI for k capacity queues in tandem ($k \in \mathbb{Z}^+$) with equal service rate μ. The results are shown in Figures 13A–D after 1–4 queues, with k = 1, 5, 10, ∞ in each plot. We observe that for finite capacity queues, the average age approaches some finite value as the arrival rate increases, and that the difference between this value and the minimum is increasingly greater for larger capacity queues. Thus, optimizing the arrival rate becomes critical very quickly, even for relatively small capacity queues.

The optimal arrival rate vs. queue capacity is shown in Figure 14 for queue capacity up to six. We observe that the optimal λ very quickly approaches the optimal for infinite capacity queues (dashed line). We have also plotted the optimal age in Figure 15 and again the optimal age for infinite capacity queues is quickly approached. These results suggest that the optimal age is very sensitive to

queue capacity between the values of one and six. Meanwhile, for larger queue capacities, it suffices to optimize for infinite capacity queues. This can be easier to analyze theoretically, such as in the case for two queues in tandem. We have simulated this for different service rates $\mu = 0.1$ and $\mu = 10$, and have observed very similar behavior. Consequently, this result is insensitive to service rate.

6 Summary

In this work, we have analyzed the average AoI for non-preemptive queues in tandem, which is important for understanding how to control and optimize multi-hop networks for information freshness. For systems with Poisson arrivals and exponential service times, we applied the stochastic hybrid systems modeling approach to derive closed form expressions for the average age after two queues in tandem with capacity 1 and different service rates, as well as for the average age after three queues in tandem with equal service rates. Based on the average age expressions for one and two queues in tandem with equal service rates, we consider an approximation for m queues in tandem, which is shown to overestimate the age for the $m = 3$ and $m = 4$ cases with a gap that increases with λ. We have also derived the closed form expression for the average age after two infinite capacity queues with equal service rates. Finally, we studied other queue capacities k besides one and infinity. We observed that even for very small values of k, the optimal λ and age are close to that of the infinite capacity queue case.

Data availability statement

The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Author contributions

CK and SK contributed to conception and design of the study. CK conducted the analysis, simulations, and wrote the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was funded by the Office of Naval Research.

Acknowledgments

The content of this manuscript has been presented in part at MILCOM 2018 (Kam et al., 2018).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abd-Elmagid, M. A., and Dhillon, H. S. (2022a). Closed-form characterization of the mgf of aoi in energy harvesting status update systems. *IEEE Transactions on Information Theory* 68, 3896–3919. doi:10.1109/TIT.2022.3149450

Abd-Elmagid, M. A., and Dhillon, H. S. (2022b). Distribution of aoi in eh-powered multi-source systems under non-preemptive and preemptive policies. in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 1–8. doi:10.1109/INFOCOMWKSHPS54753.2022.9798131

Bedewy, A. M., Sun, Y., and Shroff, N. B. (2017). "Age-optimal information updates in multihop networks," in 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, June 25 to 30, 2017, 576–580. doi:10.1109/ISIT.2017.8006593

Bedewy, A. M., Sun, Y., and Shroff, N. B. (2016). "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10-15 July 2016, 2569–2573. doi:10.1109/ISIT.2016.7541763

Burke, P. J. (1956). The output of a queuing system. *Operations Res.* 4, 699–704. doi:10.1287/opre.4.6.699

Chiariotti, F., Vikhrova, O., Soret, B., and Popovski, P. (2020). Peak age of information distribution in tandem queue systems. arXiv e-prints , arXiv–2004. Available at: https://arxiv.org/pdf/2004.05088.pdf.

Costa, M., Codreanu, M., and Ephremides, A. (2016). On the age of information in status update systems with packet management. *IEEE Trans. Inf. Theory* 62, 1897–1910. doi:10.1109/TIT.2016.2533395

Kam, C., Kompella, S., and Ephremides, A. (2013). "Age of information under random updates," in Proc. IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 7-12 July 2013, 66–70.

Kam, C., Kompella, S., Nguyen, G. D., and Ephremides, A. (2016a). Effect of message transmission path diversity on status age. *IEEE Trans. Inf. Theory* 62, 1360–1374. doi:10.1109/TIT.2015.2511791

Kam, C., Kompella, S., Nguyen, G. D., Wieselthier, J. E., and Ephremides, A. (2016b). "Controlling the age of information: Buffer size, deadline, and packet replacement," in 2016 Military Communications Conference (MILCOM 2016), Baltimore, MD, USA, 1-3 Nov. 2016.

Kam, C., Molnar, J. P., and Kompella, S. (2018). "Age of information for queues in tandem," in MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM), Angeles, CA, USA, 29-31 Oct. 2018. 1–6. doi:10.1109/MILCOM.2018.8599728

Karpelevitch, F., and Kreinin, A. Y. (1992). Joint distributions in poissonian tandem queues. *Queueing Syst.* 12, 273–286. doi:10.1007/bf01158803

Kaul, S., Gruteser, M., Rai, V., and Kenney, J. (2011). "Minimizing age of information in vehicular networks," in IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Utah, USA, June 27-30,2011, 350–358.

Kaul, S., Yates, R., and Gruteser, M. (2012a). "Real-time status: How often should one update?," in Proc. IEEE INFOCOM, Orlando, FL, 25-30 March 2012, 2731–2735.

Kaul, S., Yates, R., and Gruteser, M. (2012b). "Status updates through queues," in Conference on Information Sciences and Systems (CISS), Princeton, NJ, 21-23 March 2012, 1–6.

Kosta, A., Pappas, N., and Angelakis, V. (2017). Age of information: A new concept, metric, and tool. *FNT. Netw.* 12, 162–259. doi:10.1561/1300000060

Koukoutsidis, I. (2020). "Age of information in an overtake- free network of quasi - reversible queues," in 2020 28th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Nice, France, November 17–19, 2020. 1–6. doi:10.1109/MASCOTS50786.2020.9285958

Kuang, Q., Gong, J., Chen, X., and Ma, X. (2019). "Age-of- information for computation- intensive messages in mobile edge computing," in 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, October 23-25,2019. 1–6. doi:10.1109/WCSP.2019.8927944

Moltafet, M., Leinonen, M., and Codreanu, M. (2021). Moment generating function of the aoi in a two-source system with packet management. *IEEE Wireless Communications Letters* 10, 882–886. doi:10.1109/LWC.2020.3048628

Soret, B., Ravikanti, S., and Popovski, P. (2020). "Latency and timeliness in multi-hop satellite networks," in ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7-11 June 2020. 1–6. doi:10.1109/ICC40277.2020.9149009

Sun, Y., Uysal-Biyikoglu, E., Yates, R. D., Koksal, C. E., and Shroff, N. B. (2017). Update or wait: How to keep your data fresh. *IEEE Trans. Inf. Theory* 63, 7492–7508. doi:10.1109/TIT.2017.2735804

Talak, R., Karaman, S., and Modiano, E. (2017). "Minimizing age-of-information in multi-hop wireless networks," in 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 3-6 Oct. 2017, 486–493. doi:10.1109/ALLERTON.2017.8262777

Vikhrova, O., Chiariotti, F., Soret, B., Araniti, G., Molinaro, A., and Popovski, P. (2020). "Age of information in multi-hop networks with priorities," in GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei City Taiwan, 7 - 11 December 2020. 1–6. doi:10.1109/GLOBECOM42002.2020.9348175

Xu, C., Yang, H. H., Wang, X., and Quek, T. Q. S. (2020). Optimizing information freshness in computing-enabled IoT networks. *IEEE Internet Things J.* 7, 971–985. doi:10.1109/JIOT.2019.2947419

Yates, R. D. (2018). "Age of information in a network of preemptive servers," in IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI, USA, 15-19 April 2018, 118–123. doi:10.1109/INFCOMW.2018.8406966

Yates, R. D., and Kaul, S. (2012). "Real-time status updating: Multiple sources," in Proc. IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, 1-6 July 2012, 2666–2670.

Yates, R. D., and Kaul, S. K. (2019). The age of information: Real-time status updating by multiple sources. *IEEE Trans. Inf. Theory* 65, 1807–1827. doi:10.1109/TIT.2018.2871079

Yates, R. D. (2020). The age of information in networks: Moments, distributions, and sampling. *IEEE Transactions on Information Theory* 66, 5712–5728. doi:10.1109/TIT.2020.2998100

Appendix: Proof of Theorem 2

Similar to the two-queue case, the states $0q_2q_3$ all have outgoing transitions through arrivals into queue 1 of rate λ, and there is an outgoing transition of rate μ for each occupied queue. For incoming transitions, the states $1q_2q_3$ have a transition of rate λ from $0q_2q_3$ and there is a transition of rate μ from a state where $q_i = 1$ to a state where $q_i = 0$ and $q_{i+1} = 1$ or if i = 3, so that the queue is followed by the monitor. The resulting global balance equations are given as follows:

$$\begin{aligned}
\lambda\bar{\pi}_{000} &= \mu\bar{\pi}_{001}\\
(\lambda+\mu)\bar{\pi}_{001} &= \mu(\bar{\pi}_{010}+\bar{\pi}_{011})\\
(\lambda+\mu)\bar{\pi}_{010} &= \mu(\bar{\pi}_{011}+\bar{\pi}_{100}+\bar{\pi}_{110})\\
(\lambda+2\mu)\bar{\pi}_{011} &= \mu(\bar{\pi}_{101}+\bar{\pi}_{111})\\
\mu\bar{\pi}_{100} &= \lambda\bar{\pi}_{000}+\mu\bar{\pi}_{101}\\
2\mu\bar{\pi}_{101} &= \lambda\bar{\pi}_{001}+\mu(\bar{\pi}_{110}+\bar{\pi}_{111})\\
2\mu\bar{\pi}_{110} &= \lambda\bar{\pi}_{010}+\mu\bar{\pi}_{111}\\
3\mu\bar{\pi}_{111} &= \lambda\bar{\pi}_{011}
\end{aligned}$$

with the solution

$$\begin{aligned}
\bar{\pi}_{000} &= \frac{\mu^3}{(\lambda+\mu)^3}, & \bar{\pi}_{001} &= \frac{\lambda\mu^2}{(\lambda+\mu)^3},\\
\bar{\pi}_{010} &= \frac{\lambda\mu(5\lambda+8\mu)}{8(\lambda+\mu)^3}, & \bar{\pi}_{011} &= \frac{3\lambda^2\mu}{8(\lambda+\mu)^3},\\
\bar{\pi}_{100} &= \frac{\lambda(\lambda^2+3\lambda\mu+4\mu^2)}{4(\lambda+\mu)^3}, & \bar{\pi}_{101} &= \frac{\lambda^2(\lambda+3\mu)}{4(\lambda+\mu)^3},\\
\bar{\pi}_{110} &= \frac{\lambda^2(3\lambda+4\mu)}{8(\lambda+\mu)^3}, & \bar{\pi}_{111} &= \frac{\lambda^3}{8(\lambda+\mu)^3}.
\end{aligned}$$

Similar to the global balance equations, the steady-state correlation between discrete and age states satisfy (6), which for the age at the monitor $\bar{\nu}_{q4}$ gives us the following equations:

$$\begin{aligned}
\lambda\bar{\nu}_{0004} &= \pi_{000}+\mu\bar{\nu}_{0013}\\
(\lambda+\mu)\bar{\nu}_{0014} &= \pi_{001}+\mu(\bar{\nu}_{0104}+\bar{\nu}_{0114})\\
(\lambda+\mu)\bar{\nu}_{0104} &= \pi_{010}+\mu(\bar{\nu}_{1004}+\bar{\nu}_{1104}+\bar{\nu}_{0113})\\
(\lambda+2\mu)\bar{\nu}_{0114} &= \pi_{011}+\mu(\bar{\nu}_{1014}+\bar{\nu}_{1114})\\
\mu\bar{\nu}_{1004} &= \pi_{100}+\lambda\bar{\nu}_{0004}+\mu\bar{\nu}_{1013}\\
2\mu\bar{\nu}_{1014} &= \pi_{101}+\lambda\bar{\nu}_{0014}+\mu(\bar{\nu}_{1104}+\bar{\nu}_{1114})\\
2\mu\bar{\nu}_{1104} &= \pi_{110}+\lambda\bar{\nu}_{0104}+\mu\bar{\nu}_{1113}\\
3\mu\bar{\nu}_{1114} &= \pi_{111}+\lambda\bar{\nu}_{0114}.
\end{aligned}$$

Using a similar approach as in the two-queue case, we take linear combinations of these equations to yield $\sum_{\bar{q}\in\mathcal{Q}}\bar{\nu}_{\bar{q}4}$ when we move all $\bar{\nu}_{\bar{q}4}$ terms to the left-hand side, such that the coefficients must satisfy

$$\begin{aligned}
a\lambda - e\lambda &= 1\\
b(\lambda+\mu) - f\lambda &= 1\\
c(\lambda+\mu) - b\mu - g\lambda &= 1\\
d(\lambda+2\mu) - b\mu - h\lambda &= 1\\
e\mu - c\mu &= 1\\
2f\mu - d\mu &= 1\\
2g\mu - c\mu - f\mu &= 1\\
3h\mu - d\mu - f\mu &= 1,
\end{aligned}$$

which is achieved with solution

$$\begin{aligned}
a &= \frac{1}{\lambda}+\frac{3}{\mu}, & b &= \frac{1}{\mu}, & c &= \frac{2}{\mu}, & d &= \frac{1}{\mu}\\
e &= \frac{3}{\mu}, & f &= \frac{1}{\mu}, & g &= \frac{2}{\mu}, & h &= \frac{1}{\mu}.
\end{aligned}$$

The linear combination of equations yields the following:

$$\begin{aligned}
\sum_{\bar{q}\in\mathcal{Q}}\bar{\nu}_{\bar{q}4} &= \left(\frac{1}{\lambda}+\frac{3}{\mu}\right)\pi_{000}+\frac{1}{\mu}\pi_{001}+\frac{2}{\mu}\pi_{010}+\frac{1}{\mu}\pi_{011}+\frac{2}{\mu}\pi_{100}+\frac{1}{\mu}\pi_{101}\\
&\quad+\frac{2}{\mu}\pi_{110}+\frac{1}{\mu}\pi_{111}+\left(\frac{\mu}{\lambda}+3\right)\bar{\nu}_{0013}+2\bar{\nu}_{0113}+3\bar{\nu}_{1013}\\
&\quad+2\bar{\nu}_{1113}
\end{aligned}$$

Next we need to find $(\frac{\mu}{\lambda}+3)\bar{\nu}_{0013}+2\bar{\nu}_{0113}+3\bar{\nu}_{1013}+2\bar{\nu}_{1113}$, which is a function of the correlation functions for the age at the third queue $\bar{\nu}_{q3}$. To find this expression, we take linear combinations of the corresponding equations shown here:

$$\begin{aligned}
(\lambda+\mu)\nu_{0013} &= \pi_{001}+\mu(\nu_{0102}+\nu_{0113})\\
(\lambda+2\mu)\nu_{0113} &= \pi_{011}+\mu(\nu_{1013}+\nu_{1113})\\
2\mu\nu_{1013} &= \pi_{101}+\lambda\nu_{0013}+\mu(\nu_{1102}+\nu_{1113})\\
3\mu\nu_{1113} &= \pi_{111}+\lambda\nu_{0113}
\end{aligned}$$

such that their coefficients satisfy

$$\begin{aligned}
a'(\lambda+\mu) - c'\lambda &= \left(\frac{\mu}{\lambda}+3\right)\\
b'(\lambda+2\mu) - a'\mu - d'\lambda &= 2\\
c'2\mu - b'\mu &= 3\\
d'3\mu - b'\mu - c'\mu &= 2.
\end{aligned}$$

The solution for the coefficients is given by

$$\begin{aligned}
a' &= \frac{8\lambda^3+33\lambda^2\mu+39\lambda\mu^2+12\mu^3}{3\lambda\mu(\lambda+2\mu)^2}\\
b' &= \frac{7\lambda+12\mu}{3\mu(\lambda+4\mu)}+\frac{2\mu}{\lambda+4\mu}a'\\
c' &= \frac{8\lambda+24\mu}{3\mu(\lambda+4\mu)}+\frac{\mu}{\lambda+4\mu}a'\\
d' &= \frac{7\lambda+20\mu}{3\mu(\lambda+4\mu)}+\frac{\mu}{\lambda+4\mu}a'.
\end{aligned}$$

Next we find $a'\mu\bar{\nu}_{0102}+c'\mu\bar{\nu}_{1102}$ *via* linear combinations of

$$\begin{aligned}
(\lambda+\mu)\nu_{0102} &= \pi_{010}+\mu(\nu_{1001}+\nu_{1102}+\nu_{0112})\\
(\lambda+2\mu)\nu_{0112} &= \pi_{011}+\mu(\nu_{1011}+\nu_{1112})\\
2\mu\nu_{1102} &= \pi_{110}+\lambda\nu_{0102}+\mu\nu_{1112}\\
3\mu\nu_{1112} &= \pi_{111}+\lambda\nu_{0112}.
\end{aligned}$$

This is achieved with coefficients that satisfy

$$\begin{aligned}
a''(\lambda+\mu) - c''\lambda &= a'\mu\\
b''(\lambda+2\mu) - a''\mu - d''\lambda &= 0\\
c''2\mu - a''\mu &= c'\mu\\
d''3\mu - b''\mu - c''\mu &= 0,
\end{aligned}$$

which has solution

$$
\begin{aligned}
a'' &= \frac{8\lambda^2 + 24\lambda\mu}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{3\lambda\mu + 8\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a' \\
b'' &= \frac{4\lambda}{3\mu(\lambda+2\mu)} + \frac{\mu}{\lambda+2\mu}a' \\
c'' &= \frac{8\lambda^2 + 32\lambda\mu + 24\mu^2}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{2\lambda\mu + 5\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a' \\
d'' &= \frac{4\lambda^2 + 16\lambda\mu + 8\mu^2}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{\lambda\mu + 3\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a'.
\end{aligned}
$$

The last set of equations used to find $a''\mu\bar{\nu}_{1001} + b''\mu\bar{\nu}_{1011}$ is given by

$$
\begin{aligned}
\mu\nu_{1001} &= \pi_{100} + \mu\nu_{1011} \\
2\mu\nu_{1011} &= \pi_{101} + \mu(\nu_{1101} + \nu_{1111}) \\
2\mu\nu_{1101} &= \pi_{110} + \mu\nu_{1111} \\
3\mu\nu_{1111} &= \pi_{111},
\end{aligned}
$$

and the coefficients must satisfy

$$
\begin{aligned}
a'''\mu &= a''\mu \\
b'''2\mu - a'''\mu &= b''\mu \\
c'''2\mu - b'''\mu &= 0 \\
d'''3\mu - b'''\mu - c'''\mu &= 0,
\end{aligned}
$$

which has solution

$$
\begin{aligned}
a''' &= a'' \\
&= \frac{8\lambda^2 + 24\lambda\mu}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{3\lambda\mu + 8\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a' \\
b''' &= \frac{a'' + b''}{2} \\
&= \frac{6\lambda^2 + 20\lambda\mu}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{2\lambda\mu + 6\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a' \\
c''' &= \frac{a'' + b''}{4} \\
&= \frac{3\lambda^2 + 10\lambda\mu}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{\lambda\mu + 3\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a' \\
d''' &= \frac{a'' + b''}{4} \\
&= \frac{3\lambda^2 + 10\lambda\mu}{3\mu(\lambda+2\mu)(\lambda+4\mu)} + \frac{\lambda\mu + 3\mu^2}{(\lambda+2\mu)(\lambda+4\mu)}a'.
\end{aligned}
$$

We substitute the values into

$$
\begin{aligned}
\sum_{\bar{q}\in\mathcal{Q}} \bar{\nu}_{\bar{q}4} &= \left(\frac{1}{\lambda} + \frac{3}{\mu}\right)\pi_{000} + \left(\frac{1}{\mu} + a'\right)\pi_{001} + \left(\frac{2}{\mu} + a''\right)\pi_{010} \\
&+ \left(\frac{1}{\mu} + b' + b''\right)\pi_{011} + \left(\frac{3}{\mu} + a'''\right)\pi_{100} + \left(\frac{1}{\mu} + c' + b'''\right)\pi_{101} \\
&+ \left(\frac{2}{\mu} + c'' + c'''\right)\pi_{110} + \left(\frac{1}{\mu} + d' + d'' + d'''\right)\pi_{111}
\end{aligned}
$$

which, after much simplification, yields the final result.