Content Based Status Updates

Elie Najm  
LTHI, EPFL, Lausanne, Switzerland  
Email: elie.najm@epfl.ch

Rajai Nasser  
LTHI, EPFL, Lausanne, Switzerland  
Email: rajai.nasser@epfl.ch

Emre Telatar  
LTHI, EPFL, Lausanne, Switzerland  
Email: emre.telatar@epfl.ch

Abstract—Consider a stream of status updates generated by a source, where each update is of one of two types: priority or ordinary; these updates are to be transmitted through a network to a monitor. We analyze a transmission policy that treats updates depending on their content: ordinary updates are served in a first-come first-served fashion, whereas the priority updates receive preferential treatment. An arriving priority update discards and replaces any currently-in-service priority update, and preempts (with eventual resume) any ordinary update. We model the arrival processes of the two kinds of updates as independent Poisson processes and the service times as two (possibly different) rate exponentials. We find the arrival and service rates under which the system is stable and give closed-form expressions for average peak age and a lower bound on the average age of the ordinary stream. We give numerical results on the average age of both streams and observe the effect of each stream on the age of the other.

I. INTRODUCTION

While the classical notion of delay is a measure of how long a packet spends in transit, the ‘Age of Information’ (AOI) is a receiver-centric notion that measures how fresh the data is at the receiver. Specifically, with \( u(t) \) denoting the generation time of the last successfully received packet before time \( t \), one defines \( \Delta(t) = t - u(t) \) as the instantaneous age of the information at the receiver at time \( t \). One can then consider

\[
\Delta = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \Delta(t) dt, \tag{1}
\]

as the (time) average age. Observe that \( \Delta(t) \) increases linearly in the intervals between packet receptions, and when a packet is received, \( \Delta(t) \) jumps down to the delay experienced by this packet. This results in a sawtooth sample path as in Fig. 1. In [1]–[6] the properties of \( \Delta \) were investigated under the assumption that the packets are generated by a Poisson process, and various transmission policies (M/M/1, M/M/\( \infty \), gamma service time,...).

A related metric, called average peak age, was introduced in [3] as the average of the value of the instantaneous age \( \Delta(t) \) at times just before its downward jumps. In Fig. 1, \( K_j \) denotes the instantaneous age just before the reception of the \( j^{th} \) successfully transmitted packet, and hence, the average peak age is given by

\[
\Delta_{\text{peak}} = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} K_j. \tag{2}
\]

The authors in [7] studied the average age when considering multiple sources sending update through one queue. They computed the average age for three scenarios: all sources transmit according to an M/M/1 FCFS policy, all sources transmit according to an M/M/1/1 with preemption policy and all sources transmit according to an M/M/1/1 with preemption in waiting policy. In the M/M/1/1 with preemption policy, if a newly generated update finds the system busy, the transmitter preempts the one currently in service and starts sending the new packet. On the other hand, in the M/M/1/1 with preemption in waiting policy, the system has a buffer of size 1 and if the generated update finds the system busy, it replaces any waiting update in the buffer. In [8], the authors also consider multiple sources transmitting through a single queue but in this case they assume a generally distributed service time. Moreover, they study two scenarios: all sources transmit according to an M/G/1 FCFS policy or all sources transmit according to an M/G/1/1 with blocking policy. For each one of these policies, the authors give the expression of the average peak age relative to each source.

In this paper, we assume updates are generated according to a Poisson process with rate \( \lambda \), and that the updates belong to two different streams where each stream \( i \) is chosen independently with probability \( p_i \), \( i = 1, 2 \). So we have two independent Poisson streams with rates \( \lambda_1 = \lambda p_1 \) and \( \lambda_2 = \lambda p_2 \). However, unlike [7] and [8], we assume a different transmission policy for each stream. To the best of our knowledge, this model was not studied before although it models a natural scenario. In fact, the two independent streams generated by the source can be used to model different types of content carried by the packets of each stream. For example, if the source is a sensor, one stream could carry emergency messages (fire alarm, high pressure, etc.) and thus it needs to be always as fresh as possible while the other stream will carry regular updates and hence is not age sensitive. Therefore, it stands to reason to transmit these two streams in a different manner. The regular stream will be transmitted according to a FCFS policy while the high priority stream will be sent by preemption, packets of the high priority stream preempt all packets including packets of their own stream. We will further assume that the service times requirements of the two streams may be different; a packet of the regular stream will be served at rate \( \mu_1 \), a packet of the priority scheme at rate \( \mu_2 \).

We will study the above model and answer the questions: what should the relation between \( \lambda_1, \mu_1, \lambda_2 \) and \( \mu_2 \) be for the system to be stable? How does each stream affects the average age of the other one? What are the ages of each stream? To answer these questions, we will give a necessary and
sufficient condition for the system stability, find the steady-state distribution of the underlying state-space, and give closed form expressions for the average peak age and a lower bound on the average age of the regular stream and compare them to the average age of the high priority stream.

This paper is structured as follows: in Section II we start by defining the model and the different variables needed in our study. In Section III we derive the stability condition of the system and its stationary distribution. The closed form expressions of the average peak age and the lower bound on the average age of the regular stream are computed in Section IV. Finally, in Section V we present numerical results of our results.

II. SYSTEM MODEL

We consider a source that generates packets (or updates) according to a Poisson process of rate $\lambda$. Each packet, independently of the previous packets, is of type 1 with probability $p_1$ and of type 2 with probability $p_2 = 1 - p_1$. We can thus see our source as generating two independent Poisson streams $U_1$ and $U_2$ with rates $\lambda_1 = \lambda p_1$ and $\lambda_2 = \lambda p_2$ respectively, $\lambda = \lambda_1 + \lambda_2$ (see [9]). As noted in the introduction, the different streams can be used to model packets of different types of content, for example, emergency messages, alerts, error messages, warnings, notices, etc.

We also assume that the updates are sent through a single server (or transmitter) queue to a monitor. The service time of each packet is considered to be exponentially distributed with rate $\mu_1$ for stream $U_1$ and rate $\mu_2$ for stream $U_2$. The difference in service rates between the two streams is to account for the possible difference in compression, packet length, etc., between the two streams.

Given this model, we impose on the transmitter that all packets from stream $U_1$ should be sent. Hence the server applies a FCFS policy on the packets from stream $U_1$ with a buffer to save waiting updates. On the other hand, we assume that the information carried by stream $U_2$ is more time sensitive (or has higher priority) and thus we aim at minimizing its average age. To this end, the transmitter is allowed to perform packet management: in this case we assume the server applies a preemption policy whenever a packet from $U_2$ is generated. This means that if a newly generated packet from stream $U_2$ finds the system busy (serving a packet from $U_1$ or $U_2$), the server preempts the update currently in service and starts serving the new packet. Moreover, if the preempted packet belongs to $U_1$, this packet is placed back at the head of the $U_1$-buffer so that it can be served once the system is idle again. If the preempted packet belongs to $U_2$ then it is discarded. However, if a newly generated $U_1$-packet finds the system busy serving a $U_2$-packet, it is placed in the buffer and served when the system becomes idle. This choice of policy for the age sensitive stream is motivated by the conclusion reached in [10] that for exponentially distributed packet transmission times, the M/M/1/1 with preemption policy is the optimal policy among causal policies.

These ideas are illustrated in part in Fig. 1 which also shows the variation of the instantaneous age of stream $U_1$. In this plot, $t_i$ and $D_i$ refer to the generation and delivery times of the $i$th packet of stream $U_1$ while $t'_i$ and $D'_i$ are the start and end times of the $i$th period during which the system is busy serving packets from stream $U_2$ only.

III. SYSTEM STABILITY AND STATIONARY DISTRIBUTION

The fact that we wish to receive all of stream $U_1$ updates and that stream $U_2$ has higher priority and preempts stream $U_1$ might lead to an unstable system. In order to derive the necessary and sufficient condition for the stability of the system we study the Markov chain of the number of packets in the system (in service and waiting) shown in Fig. 2. In this chain, $q_0$ is the idle state where the system is completely empty. States $q_i$, $i > 0$, in the upper row refer to states where the queue is serving a packet from stream $U_1$ while states $q'_i$, $i > 0$, in the row below correspond to the queue serving a packet from stream $U_2$. In both cases there are $i - 1$ stream $U_1$ updates waiting in the buffer.

The system leaves state $q_0$ at rate $\lambda_1$ to state $q_1$ when a packet from stream $U_1$ is generated first and it leaves $q_0$ at rate $\lambda_2$ to state $q'_1$ when a packet from stream $U_2$ is generated first. However, when the system enters state $q_i$, $i > 0$, three exponential clocks start: a clock with rate $\mu_1$ which corresponds to the service time of the stream $U_1$ packet being served, a clock with rate $\lambda_1$ which corresponds to the generation time of stream $U_1$ packets and a clock with rate $\lambda_2$ which corresponds to the generation time of stream $U_2$ packets. If the $\mu_1$-clock ticks first, the system goes to state $q_{i-1}$; this means that the current stream $U_1$ packet was delivered and the queue begins the service of the next one in the buffer (if there is any). However, if the $\lambda_1$-clock ticks first, a new stream $U_1$ update is generated and added to the buffer and hence the system goes to state $q_{i+1}$. On the other hand, if the $\lambda_2$-clock ticks first, the system preempts the packet currently in service and places it back at the head of the buffer and starts the service of the newly generated stream $U_2$ update.
Thus the system goes to state \( q'_{i+1} \). When the system enters a state \( q'_{i} \), \( i > 0 \), two exponential clocks start: the clock with rate \( \lambda_1 \) and a clock with rate \( \mu_2 \) which corresponds to the service time of a stream \( U_2 \) packet. If the \( \lambda_1 \)-clock ticks first, the newly generated stream \( U_1 \) packet is placed in the buffer and the stream \( U_2 \) update is continued to be served. Hence the system goes to state \( q'_{i+1} \). However, if the \( \mu_2 \)-clock ticks first, the stream \( U_2 \) packet has finished service and the system starts serving the first stream \( U_1 \) packet in the buffer (if there is any). Thus the system goes to state \( q_{i-1} \).

This next theorem gives the necessary and sufficient condition for the above system to be stable as well as its stationary distribution.

**Theorem 1.** The system described in Section II is stable, i.e. the average number of packets in the queue is finite, if and only if

\[
\mu_1 > \lambda_1 \left(1 + \frac{\lambda_2}{\mu_2}\right).	ag{3}
\]

In this case the Markov chain shown in Fig. 2 has a stationary distribution \( \Pi = [\pi_0, \pi_1, \ldots, \pi_i, \ldots, \pi_{q_1'}, \ldots, \pi_{q_2'}, \ldots] \), where \( \pi_i \) denotes the stationary probability of state \( q_i \), \( i \geq 0 \), and \( \pi_{q_i'} \) denotes the stationary probability of state \( q'_{i} \), \( i > 0 \). This stationary distribution is described by the following system of equations,

\[
\begin{align*}
\pi_0 &= \frac{\mu_2}{\mu_2 + \lambda_2} - \frac{\lambda_1}{\mu_1}, \\
\begin{bmatrix} \pi_i \\ \pi_{q_i'} \end{bmatrix} &= \begin{bmatrix} \frac{\lambda_1}{\mu_1} - \frac{\mu_2 \lambda_2}{\lambda_2 + \mu_2} \\ \frac{\mu_2 \lambda_2}{\lambda_2 + \mu_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \pi_0, 
\end{align*}
\]

where \( \lambda = \lambda_1 + \lambda_2 \), \( H = \begin{bmatrix} C & D \\ I_2 & 0 \end{bmatrix} \),

\[
C = \begin{bmatrix} 1 + \frac{\lambda_1}{\mu_1} - \frac{\mu_2 \lambda_2}{\mu_2 + \lambda_1} & -\frac{\mu_2 \lambda_2}{\mu_2 + \lambda_1} \\ \frac{\mu_2 \lambda_2}{\mu_2 + \lambda_1} & \frac{\lambda_1}{\mu_1} \end{bmatrix}, \\
D = \begin{bmatrix} -\frac{\lambda_1}{\mu_1} \\ 0 \end{bmatrix}.
\]

\( I_2 \) is the 2 \times 2 identity matrix and \( 0 \) is the 2 \times 2 zero matrix.

**Corollary 1.** If we define \( N(t) \) to be the number of stream \( U_1 \) packets in the system at time \( t \), then its moment generating function is \( \phi_{N(t)}(s) \)

\[
\phi_{N(t)}(s) = \frac{\pi_0 \mu_1 (\lambda_1 + \lambda_2 + \mu_2 - 4s)}{\mu_1 \mu_2 + \mu_1 \lambda_1 - \mu_2 \lambda_1},
\]

where \( \pi_0 \) is given by (4). Particularly, the expected value of \( N(t) \) is

\[
\mathbb{E}(N(t)) = \frac{\lambda_1 (2\lambda_2 \mu_2 + \lambda_2 \mu_1 + \lambda_2^2 + \mu_2^2)}{(\mu_2 + \lambda_2)(\mu_1 \mu_2 - \lambda_1 (\mu_2 + \lambda_2))}.	ag{7}
\]

**Proof.** The distribution given by (4) and (5) satisfy the detailed balance equations of the Markov chain shown in Fig. 2. Moreover, (5) is the condition needed to have \( \pi_0 > 0 \). As for the expression for \( \phi_{N(t)}(s) \), it is a consequence of (4) and (5). The appendix presents a full technical version of the proof for Theorem 1 and Corollary 1.

The condition in (5) can be interpreted as follows: define the map \( f \) from the state-space of the chain as \( f(s) = 0 \) if \( s \) is in \( \{q_0, q_1, \ldots\} \) and \( f(s) = 1 \) if \( s \in \{q'_1, q'_2, \ldots\} \). For each \( s \) and \( s' \), for which \( f(s) = 0 \) and \( f(s') = 1 \) the transition rate from \( s \) to \( s' \) is the same (\( \lambda_2 \)) and similarly for \( s \) and \( s' \) with \( f(s) = 1 \), \( f(s') = 0 \), (\( \mu_2 \)). Consequently \( F(t) = f(s(t)) \), with \( s(t) \) being the state at time \( t \), is Markov (which would not be the case for an arbitrary \( F \)), and it is easily seen that \( F(t) = 0 \) a fraction \( \phi_0 = \mu_2 / (\lambda_2 + \mu_2) \) amount of time, \( F(t) = 1 \) a fraction \( \phi_1 = \lambda_2 / (\lambda_2 + \mu_2) \) amount of time. Thus, while the Markov chain in Fig. 2 moves right at rate \( \lambda_1 \), it moves left at a rate \( \mu_1 \phi_0 \). The system is stable only if the rate of moving left is larger than the rate of moving right; which gives the condition (5).

**IV. AGES OF STREAMS \( U_1 \) AND \( U_2 \)**

**A. Preliminaries**

In this section, unless stated otherwise, all random variables correspond to stream \( U_1 \). We also follow the convention where a random variable \( U \) with no subscript corresponds to the steady-state version of \( U_j \) which refers to the random variable relative to the \( j \)th received packet from stream \( U_1 \). To differentiate between streams we will use superscripts, so \( U^{(i)} \) corresponds to the steady-state variable \( U \) relative to stream \( U_i \), \( i = 1, 2 \).

In addition to that, we define: (i) \( X^{(i)} \) to be the interarrival time between two consecutive generated updates from stream \( U_i \), so \( f_{X^{(i)}}(x) = \lambda_i e^{-\lambda_i x}, i = 1, 2 \) (ii) \( S^{(i)} \) to be the service time random variable of stream \( U_i \) updates, so \( f_{S^{(i)}}(t) = \mu_i e^{-\mu_it}, i = 1, 2 \) (iii) \( T_j \) to be the system time, or the time spent by the \( j \)th stream \( U_i \) update in the queue. In our model, we assume the service time of the updates from the different streams to be independent of the interarrival time between consecutive packets (belonging to the same stream or not).

Given the description of the model in Section II, we can define for each \( U_i \) packet \( j \) a “virtual” service time \( Z_j \) which could be different from its “physical” service time \( S_j^{(i)} \). We define the “virtual” service time \( Z_j \) as follows:

\[
Z_j = D_j - \max(D_{j-1}, t_j),
\]

where \( D_j \) is the delivery time of the \( j \)th packet and \( t_j \) is its generation time. Fig. 1 shows the “virtual” service time for packets 3 and 4.
Thus, two clocks start: a service clock $S^{(2)}$ and a clock $X^{(2)}$. The service clock ticks first with probability $u = \mathbb{P}(S^{(2)} < X^{(2)})$ and its value $U$ has distribution $\mathbb{P}(U > t) = \mathbb{P}(S^{(2)} > t|S^{(2)} < X^{(2)})$. At this point, the packet currently being served finishes service before any new stream $U_2$ packet is generated and the system goes back to state $s_1$ where the $j^{th}$ packet of stream $U_1$ starts service again. However, clock $X^{(2)}$ ticks first with probability $b = 1 - u$ and its value $B = \mathbb{P}(B > t) = \mathbb{P}(X^{(2)} > t|X^{(2)} < S^{(2)})$. At this point, a new stream $U_2$ update is generated and preempts the one currently in service. In this case the system stays in state $s_2$.

From the above analysis we see that the “virtual” service time is given by the sum of the values of the different clocks on the path starting and finishing at $s_0$. For example, for the path $s_0s_1s_2s_1s_2s_2s_1s_0$ in Fig. 3, the “virtual” service time $Y = V_1 + U_1 + V_2 + B_1 + U_2 + A_1$, where all the random variables in the sum are mutually independent. This value of $Y$ is also valid for the path $s_0s_1s_2s_2s_1s_2s_1s_0$. Hence $Y$ depends on the variables $A_1, B_1, U_1, V_1$ and their number of occurrences and not on the path itself. Therefore, the probability that exactly $(i_1, i_2, i_3, i_4)$ occurrences of $(A, B, U, V)$ happen, which is equivalent to the probability that

$$Y = \sum_{k=1}^{i_1} A_k \sum_{k=1}^{i_2} B_k \sum_{k=1}^{i_3} U_k \sum_{k=1}^{i_4} V_k$$

is given by $a^{i_1} b^{i_2} u^{i_3} v^{i_4} Q(i_1, i_2, i_3, i_4)$, where $Q(i_1, i_2, i_3, i_4)$ is the number of paths with this combination of occurrences. Taking into account the fact that the sequence $A_k, B_k, U_k, V_k$ are mutually independent and denoting by $\{I_1, I_2, I_3, I_4\}$ the random variables associated with the number of occurrences of $\{A, B, U, V\}$ respectively, the moment generating function of $Y$ is

$$\phi_Y(s) = \mathbb{E}(e^{sY}) = \mathbb{E}(e^{sA^1} e^{sB^2} e^{sU^3} e^{sV^4} Q(i_1, i_2, i_3, i_4))$$

(11)

However (11) is nothing but the generating function $H_1(D_1, D_2, D_3, D_4)$ of the detour flow graph shown in Fig. 4(a) where $D_1, D_2, D_3, D_4$ are dummy variables (see [11] pp. 213–216). Simple calculations give

$$H_1(D_1, D_2, D_3, D_4) = \sum_{i_1, i_2, i_3, i_4} a^{i_1} b^{i_2} u^{i_3} v^{i_4} D_1^{i_1} D_2^{i_2} D_3^{i_3} D_4^{i_4}$$

(12)

Thus

$$\phi_Y(s) = H_1(e^{sA}, e^{sB}, e^{sU}, e^{sV})$$

For stream $U_1$, given that the average age calculations seems to be intractable, we will compute its average peak age and give a lower bound on its average age. To that end, we first study the steady state “virtual” service time $Z$.

We define the event

$$\Psi_j = \{\text{packet } j \text{ finds the system in state } q'_1\}$$

and its complement $\overline{\Psi}_j$. Then we need the following lemmas.

Lemma 1. Let $Y_j$ be the “virtual” service time of packet $j$ given that this packet does not find the system in state $q'_1$, i.e.

$$\mathbb{P}(Y_j > t) = \mathbb{P}(Z_j > t|\overline{\Psi}_j).$$

Then, in steady state,

$$\phi_{Y_j}(s) = \mathbb{E}(e^{sY_j}) = \frac{\mu_1(\mu_2 - s)}{s^2 - s(\mu_2 + \mu_1 + \lambda_2) + \mu_1 \mu_2}. \quad (9)$$

Similarly, let $Y'_j$ be the “virtual” service time of packet $j$ given that this packet finds the system in state $q'_1$, i.e.

$$\mathbb{P}(Y'_j > t) = \mathbb{P}(Z'_j > t|\Psi_j).$$

Then, in steady state,

$$\phi_{Y'_j}(s) = \mathbb{E}(e^{sY'_j}) = \frac{\mu_1 \mu_2}{s^2 - s(\mu_2 + \mu_1 + \lambda_2) + \mu_1 \mu_2}. \quad (10)$$

Proof. We start by proving (9). For that we will use the detour flow graph method.

Fig. 3 shows the semi-Markov chain relative to the “virtual” service time $Y_j$ of the $j^{th}$ packet of first stream $U_1$. When the $j^{th}$ packet gets at the head of the buffer, the system is in the idle state $s_0$. Hence with probability $1 - b$ it goes immediately to state $s_1$ where it starts serving the $j^{th}$ packet. Due to the memoryless property of the interarrival time of the second stream $X^{(2)}$, two clocks start: a service clock $S^{(1)}$ and a clock $X^{(2)}$. The service clock ticks first with probability $a = \mathbb{P}(S^{(1)} < X^{(2)})$ and its value $A$ has distribution $\mathbb{P}(A > t) = \mathbb{P}(S^{(1)} > t|S^{(1)} < X^{(2)})$. At this point the stream $U_1$ packet currently being served finishes service before any packet from the other stream is generated and the system goes back to state $s_0$. This ends the “virtual” service time $Y_j$. On the other hand, clock $X^{(2)}$ ticks first with probability $u = 1 - a = \mathbb{P}(X^{(2)} < S^{(1)})$ and its value $V$ has distribution $\mathbb{P}(V > t) = \mathbb{P}(X^{(2)} > t|X^{(2)} < S^{(1)})$. At this point, a new stream $U_2$ update is generated and preempts the stream $U_1$ packet currently in service. In this case the system goes to state $s_2$, where the preempted stream $U_1$ update is placed back at the head of the buffer and the system starts service of the stream $U_2$ update.

When the system arrives in state $s_2$, this means a new stream $U_2$ packet was just generated and is starting service.
From [7] Appendix A, Lemma 2, we know that \(A, B, U\) and \(V\) are exponentially distributed with \(E(e^{\theta A}) = E(e^{\theta U}) = \frac{\lambda_2 + \mu_2}{\lambda_2 + \mu_2 - \rho}\) and \(E(e^{\theta A}) = E(e^{\theta V}) = \frac{\lambda_2 + \mu_1}{\lambda_2 + \mu_2 - \rho}\). Simple computations show that \(a = \frac{\mu_2}{\mu_1 + \lambda_2}, b = \frac{\lambda_2}{\mu_2 + \lambda_2}, u = \frac{\mu_2}{\mu_2 + \lambda_2}, v = \frac{\lambda_2}{\mu_2 + \lambda_2}\). Finally, replacing the above expressions into (12), we get our result.

To prove (10), we use the same method as before but in this case we notice that the \(j\)th packet from stream \(U_1\) finds the system busy serving a packet from stream \(U_2\). This translates in the detour flow graph shown in Fig. 4. The generating function of this graph is

\[
H_2(D_1, D_2, D_3, D_4) = \frac{aD_1uD_3}{1 - bD_2 - vD_4uD_3}. \tag{13}
\]

For \((D_1, D_2, D_3, D_4) = (E(e^{\theta A}), E(e^{\theta B}), E(e^{\theta U}), E(e^{\theta V}))\) and replacing \(a, b, u, v\) by their values in (13), we get (10).

**Lemma 2.** The first and second moments of the “virtual” service time \(Z\) are given by

\[
E(Z) = \frac{\lambda_2}{(\lambda_1 + \mu_2)(\mu_2 + \lambda_2)} + \frac{\lambda_1 + \lambda_2 + \mu_2}{\mu_1(\lambda_1 + \mu_2)}.
\]

\[
E(Z^2) = \frac{2((\lambda_2 + \mu_2)^2(\lambda_2 + \mu_2 + \lambda_1) + \lambda_1 \mu_1(2\lambda_2 + \mu_1 + 2\mu_2))}{\mu_1^2 \mu_2(\lambda_1 + \mu_2)(\lambda_2 + \mu_2)} \tag{14}
\]

**Proof.** For any packet \(j\) of stream \(U_1\), conditioning on the event \(\Psi_j\), we get

\[
E(Z_j) = P(\Psi_j)E(Z_j | \Psi_j) + P(\Psi_j)E(Z_j | \Psi_j) = P(\Psi_j)E(Y_j') + P(\Psi_j)E(Y_j), \tag{15}
\]

where \(Y_j'\) and \(Y_j\) are defined as in Lemma 1. From Theorem 1 we deduce that \(P(\Psi_j) = \pi'_1 = \frac{\lambda_2}{\mu_2 + \mu_2 - \rho} \pi_0\). In steady-state, \(E(Z) = \pi'_1 E(Y') + (1 - \pi'_1) E(Y)\). Moreover, using (9) and (10) we get

\[
E(Y) = \frac{\mu_2 + \lambda_2}{\mu_1 \mu_2}, \quad E(Y') = \frac{\mu_1 + \mu_2 + \lambda_2}{\mu_1 \mu_2}.
\]

Similarly, \(E(Z^2) = \pi'_1 E(Y''') + (1 - \pi'_1) E(Y^2)\). Using (9) and (10) we get

\[
E(Y') = \frac{2((\mu_2 + \lambda_2)^2 + \mu_1 \lambda_2)}{(\mu_1 \mu_2)^2}
\]

\[
E(Y^2) = \frac{2((\mu_1 + \mu_2 + \lambda_2)^2 - \mu_1 \mu_2)}{(\mu_1 \mu_2)^2}.
\]

**B. Average peak age of stream \(U_1\)**

It is worth noting that the system under consideration can’t be seen as an M/G/1 queue with service time distributed as \(Z\) since the “virtual” service times of different packets are correlated. Indeed, if we know that the “virtual” service time of packet \(j\), \(Z_j\), is big then with very high probability the \((j+1)\)th packet will be generated during the service of the \(j\)th packet and thus with high probability \(Z_{j+1}\) will be distributed as \(Y\). On the other hand, if \(Z_j\) is small then there is a non-negligible probability with which the \((j+1)\)th packet will find the system serving stream \(U_2\) and thus \(Z_{j+1}\) will be distributed as \(Y'\).

**Theorem 2.** The average peak age of stream \(U_1\) is given by

\[
\Delta_{peak,1} = \frac{1}{\lambda_1} + \frac{2\lambda_2 \mu_2 + \lambda_2 \lambda_1 + \lambda_1 \lambda_2 + \mu_2^2}{(\mu_2 + \lambda_2)(\mu_1 \lambda_2 - \mu_1 \mu_2 - \lambda_1 \mu_2 - \lambda_1 \lambda_2)}. \tag{16}
\]

**Proof.** As we can deduce from Fig. 1, the \(j\)th peak \(K_j = X_j^{(1)} + T_j\) where \(X_j^{(1)}\) is the \(j\)th interarrival time for stream \(U_1\) and \(T_j\) is the service time of the \(j\)th stream \(U_1\) update. At steady state, we get \(\Delta_{peak,1} = E(K_j) = E(X_j^{(1)}) + E(T_j)\). From Little’s law we know that \(E(T_j) = E(N(t)) E(X_j^{(1)})\), with the expected number of stream \(U_1\) packets \(E(N(t))\) given by (7) and \(E(X_j^{(1)}) = 1/\lambda_1\).

**C. Lower bound on the average age of stream \(U_1\)**

The average peak age is an obvious upper bound, hence in this section we will compute a lower bound of the average age.

Consider a fictitious system where if a stream \(U_1\) arrival finds the system in state \(q_1\), then the stream \(U_1\) packet that is being served is discarded (and the stream \(U_1\) packet enters service immediately). The instantaneous age process of this fictitious system is pointwise less than the instantaneous age of the true system, consequently its average age lower bounds the true average age. Note that the fictitious system from the point of view of the stream \(U_1\) is M/G/1, with service time distributed like \(Y\) in (9).

**Lemma 3.** Assume an M/G/1 queue with interarrival time \(X_j^{(1)}\) exponentially distributed with rate \(\lambda_1\) and service time \(Y\) whose moment generating function is given by (7). The service time and the interarrival time are assumed to be independent. Then the distribution of the system time \(T\) is

\[
f_T(t) = -C_1 e^{-\alpha_1 t} - C_2 e^{-\alpha_2 t}, \quad t \geq 0, \tag{17}
\]

where \(\alpha_1 > \alpha_2 > 0\) are the roots of the quadratic expression

\[
s^2 - s(\mu_1 + \mu_2 + \lambda_2 - \lambda_1) + \mu_1 \mu_2 - \lambda_1 \mu_2 - \lambda_1 \lambda_2,
\]
\[ C_1 = \frac{(1 - \rho)\mu_1(\mu_2 - \alpha_1)}{\alpha_1 - \alpha_2}, \quad C_2 = \frac{(1 - \rho)\mu_1(\mu_2 - \alpha_2)}{\alpha_2 - \alpha_1} \]

and \( \rho = \lambda_1\mathbb{E}(Y) = \frac{\lambda_1(\mu_2 + \lambda_2)}{\mu_1}\mu_2^{\alpha_1}\mu_2^{\alpha_2}. \)

**Proof.** From [12, p. 166], we know that

\[ \phi_T(s) = \frac{(1 - \rho)\phi_Y(s)}{s + \lambda_1(1 - \phi_Y(s))}. \]

Replacing \( \phi_Y(s) \) by its expression in (9) we get

\[ \phi_T(s) = \frac{s^2 - s(\mu_1 + \mu_2 + \lambda_2 - \lambda_1) + \mu_1\mu_2 - \lambda_1\lambda_2}{s - \alpha_1 + \frac{C_1}{s - \alpha_2}} \]

by partial fraction expansion. Moreover, due to condition (3), \( \alpha_1 + \alpha_2 = \mu_1 + \mu_2 + \lambda_2 - \lambda_1 > 0 \)

and \( \alpha_1\alpha_2 = \mu_1\mu_2 - \lambda_1\mu_2 - \lambda_1\lambda_2 > 0. \)

This proves that both roots \( \alpha_1 \) and \( \alpha_2 \) are positive. Without loss of generality, we take \( \alpha_1 > \alpha_2. \)

Taking the inverse Laplace transform of \( \phi_T(-s) \) we get (17).

From [1], we know that the average age of the M/G/1 queue with interarrival time \( X^{(1)} \) and service time \( Y \) is

\[ \Delta_{LB} = \lambda_1 \left( \frac{1}{2} \mathbb{E}\left(X^{(1)}_j\right)^2 + \mathbb{E}\left(T_jX^{(1)}_j\right) \right) \]

where for the \( j \)th packet we have \( T_j = (T_j - X^{(1)}_j)^+ + Y_j, \) \( f(x) = (x)^+ = x_{1\{x>0\}} \) and \( 1\{ \cdot \} \) is the indicator function.

So \( \mathbb{E}\left(T_jX^{(1)}_j\right) \) becomes

\[ \mathbb{E}\left(T_jX^{(1)}_j\right) = \mathbb{E}\left(X^{(1)}_j(T_j - X^{(1)}_j)^+\right) + \mathbb{E}(Y_j) \mathbb{E}\left(X^{(1)}_j\right) \]

where the second term is due to the independence between \( Y_j \) and \( X_j. \)

**Proposition 1.**

\[ \mathbb{E}\left(X^{(1)}_j(T_j - X^{(1)}_j)^+\right) = \frac{\lambda_1(1 - \rho)\mu_1\mu_2^2(\alpha_1^2 + 2\lambda_1\mu_2 - \alpha_1\alpha_2)}{\mu_1^2(\lambda_1 + \mu_2)^2(\alpha_1\alpha_2)^2} + \lambda_1(1 - \rho)\mu_1\mu_2^2(\lambda_1^2\alpha_2 - \lambda_1\lambda_2\alpha_2) \]

where \( \alpha_1 + \alpha_2 = \mu_1 + \mu_2 + \lambda_2 - \lambda_1 \) and \( \rho = \lambda_1\mathbb{E}(Y) = \frac{\lambda_1(\mu_2 + \lambda_2)}{\mu_1}\mu_2^{\alpha_1}\mu_2^{\alpha_2}. \)

**Proof.** Given that \( T_j \) and \( X^{(1)}_j \) are independent then

\[ \mathbb{E}\left(X^{(1)}_j(T_j - X^{(1)}_j)^+\right) = \int_0^\infty \int_0^\infty x(t-x)f_T(t)\lambda_1 e^{-\lambda_1 x} dt dx \]

Replacing \( f_T(t) \) by its value in (17), we get (21) after some computations.

Finally, using (21), \( \mathbb{E}(Y_j) = \mathbb{E}(Y) = \frac{\mu_2 + \lambda_2}{\mu_1\mu_2} \) and \( \mathbb{E}\left(X^{(1)}_j\right) = \mathbb{E}\left(X^{(1)}_j\right) = \frac{1}{\lambda_1}, \) we can find a closed form expression for \( \mathbb{E}\left(T_jX^{(1)}_j\right) \)

and using the fact that \( \mathbb{E}\left(X^{(1)}_j\right)^2 = \frac{2}{\lambda_1^2}, \) we obtain a closed form expression of the average age \( \Delta_{LB} \) of an M/G/1 queue with interarrival time \( X^{(1)} \) and service time \( Y. \)

This is also a lower bound on the true average age of stream \( U_1. \)

**D. Average age of stream \( U_2. \)**

By design, stream \( U_2 \) is not interfered at all by stream \( U_1 \) and hence behaves like a traditional M/M/1/1 with preemption queue with generation rate \( \lambda_2 \) and service rate \( \mu_2. \)

The average age of this stream was computed in [2] to be

\[ \Delta_{U_2} = \frac{1}{\mu_2} + \frac{1}{\lambda_2}. \]

**V. Numerical results.**

Fig. 5 shows the simulated average age, the average peak age (\( \Delta_{peak,1} \)) and the lower bound on the average age (\( \Delta_{LB} \)) as computed in the previous section for stream \( U_1 \) as well as the average age (\( \Delta_{U_2} \)) of stream \( U_2. \)

In this plot, we fix \( \mu_1 = 10, \mu_2 = 5, \lambda_1 = 2 \) and we vary \( \lambda_2. \)

As we can see, for stream \( U_1 \) the average age, the lower bound and the average peak age blow up when \( \lambda_2 \) gets close to \( \frac{\mu_2}{\mu_1} \).

This observation is in line with our result in Theorem 1 and the stability condition (3). In this simulation we also notice that the average peak age appears to be a good upper bound on the average age while the lower bound is much looser.

It is easy to see via a coupling argument that if we increase \( \lambda_2, \) the age process \( \Delta_{U_1}(t) \) of the \( U_1 \) stream will stochastically increase. We see from the plots that the lower bound on \( \Delta_{U_1} \) and its average peak exhibit the same behavior. On the other hand, the average age of stream \( U_2 \) is decreasing in \( \lambda_2 \) (from (22)). Consequently, minimizing \( \Delta_{U_2} \) and minimizing \( \Delta_{U_1} \) are conflicting goals.

We have seen that the average age of stream \( U_2 \) is not affected by the presence of the other stream. However, Fig. 5
shows the effect of stream $U_2$ on the average age of stream $U_1$ ($\Delta_1$). For that, we plot the average age ($\Delta_{r-c}$) of an M/M/1 queue with generation rate $\lambda_1 = 2$ and service rate $\mu_1 = 10$ (given in [1]). We observe an expected behavior: for very low values of $\lambda_2$, the two average ages are close (they are equal at $\lambda_2 = 0$). However, as $\lambda_2$ increases the presence of stream $U_2$ quickly leads to an increase in $\Delta_1$. In fact, for $\lambda_2 = 5$, $\Delta_1$ is already 50% higher than $\Delta_{r-c}$. This shows that the presence of the priority stream $U_2$ takes a big toll on the stream $U_1$ age. Another observation is that the average age curve of stream $U_2$ crosses the average age of stream $U_1$ at a value of $\lambda_2$, denoted $\lambda_2^* = 1.9$. This means that for $\lambda_2 \leq \lambda_2^*$, stream $U_2$ has a higher average age than stream $U_1$. These observations show that not all values of $\lambda_2$ are suitable for our system. A small $\lambda_2$ will not ensure for stream $U_2$ the priority it needs while a large $\lambda_2$ will make the average age of stream $U_1$ large and the system unstable.

VI. Conclusion

In this paper we studied the effect of implementing content-dependent policies on the average age of the packets. We considered a source generating two independent Poisson streams with one stream being age sensitive and having higher priority than the other stream. The “high priority” stream is sent using a preemption policy while the “regular” stream is transmitted using a First Come First Served (FCFS) policy. We derived the stability condition for the system as well as closed form expressions for the average peak age and a lower bound on the average age of the “regular” stream. We also deduced that one can’t hope to minimize both streams if we can only control the generation rate of the high priority stream.

Acknowledgements

This research was supported in part by grant No. 200021_166106/1 of the Swiss National Science Foundation.

References

[1] S. Kaul, R. D. Yates, and M. Gruteser, “Real-time status: How often should one update?” in Proc. INFOCOM, 2012.
[2] ———, “Status updates through queues,” in Conf. on Information Sciences and Systems (CISS), Mar. 2012.
[3] M. Costa, M. Codreanu, and A. Ephremides, “Age of information with packet management,” in Proc. IEEE Int’l. Symp. Info. Theory, June 2014, pp. 1583–1587.
[4] C. Kam, S. Kompella, and A. Ephremides, “Age of information under random updates,” in Proc. IEEE Int’l. Symp. Info. Theory, 2013, pp. 669–673.
[5] E. Najm and R. Nasser, “The age of information: The gamma awakening,” in Proc. IEEE Int’l. Symp. Info. Theory, 2016, pp. 2574–2578.
[6] R. D. Yates and S. Kaul, “Real-time status updating: Multiple sources,” in Proc. IEEE Int’l. Symp. Info. Theory, Jul. 2012.
[7] R. D. Yates and S. K. Kaul, “The age of information: Real-time status updating by multiple sources,” CoRR, vol. abs/1608.08622, 2016. [Online]. Available: http://arxiv.org/abs/1608.08622
[8] L. Huang and E. Modiano, “Optimizing age-of-information in a multi-class queuing system,” in Proc. IEEE Int’l. Symp. Info. Theory, Jun. 2015.
[9] S. M. Ross, Stochastic Processes (Wiley Series in Probability and Statistics), 2nd ed. Wiley, Feb. 1995.
[10] A. M. Bedewy, Y. Sun, and N. B. Shroff, “Age-optimal information updates in multihop networks,” CoRR, vol. abs/1701.05711, 2017. [Online]. Available: http://arxiv.org/abs/1701.05711
[11] B. Rimoldi, Principles of Digital Communication: A Top-Down Approach. Cambridge University Press, 2016.
[12] J. Daigle, Queueing Theory with Applications to Packet Telecommunication, ser. Queueing Theory with Applications to Packet Telecommunication. Springer, 2005. [Online]. Available: https://books.google.com.lb/books?id=EkSsg-r-dZG6C

Appendix

Proof. (Theorem 1) Assume that
\[ \mu_1 > \lambda_1 \left( 1 + \frac{\lambda_2}{\mu_2} \right). \] (23)
The detailed balance equations of the Markov chain given by Fig. 2 are given by:
\[
\begin{align*}
\pi_{i+1} &= \left( 1 + \frac{\lambda}{\mu_1} - \frac{\mu_2 \lambda_2}{\mu_1 (\mu_2 + \lambda_1)} \right) \pi_i - \frac{\mu_2 \lambda_1}{\mu_1 (\mu_2 + \lambda_1)} \pi_i^t - \frac{\lambda_1}{\mu_1} \pi_{i-1}, \\
\pi_{i+1}^t &= \frac{\lambda_2}{\mu_2 + \lambda_1} \pi_i + \frac{\lambda_1}{\mu_2 + \lambda_1} \pi_i^t,
\end{align*}
\] (24)
where $\lambda = \lambda_1 + \lambda_2$. For easier notation we will denote
\[
\begin{align*}
a_1 &= 1 + \frac{\lambda}{\mu_1} - \frac{\mu_2 \lambda_2}{\mu_1 (\mu_2 + \lambda_1)}, \\
a_2 &= \frac{\mu_2 \lambda_1}{\mu_1 (\mu_2 + \lambda_1)}, \\
a_3 &= \frac{\lambda_1}{\mu_1}, \\
a_4 &= \frac{\lambda_2}{\mu_2 + \lambda_1}, \\
a_5 &= \frac{\lambda_1}{\mu_2 + \lambda_1}.
\end{align*}
\]
Rewriting (24) in matrix form and using the above notation, we get
\[
\begin{bmatrix}
\pi_{i+1} \\
\pi_{i+1}^t \\
\pi_i \\
\pi_i^t
\end{bmatrix} = 
\begin{bmatrix}
a_1 & -a_2 & -a_3 & 0 \\
a_4 & a_5 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\pi_i \\
\pi_i^t \\
\pi_{i-1} \\
\pi_{i-1}^t
\end{bmatrix}.
\]
Let $A_i = 
\begin{bmatrix}
\pi_{i+1} \\
\pi_{i+1}^t \\
\pi_i \\
\pi_i^t
\end{bmatrix}$, $C = \begin{bmatrix} a_1 & -a_2 \end{bmatrix}$, $D = \begin{bmatrix} -a_3 & 0 \\ 0 & 0 \end{bmatrix}$ and $H = \begin{bmatrix} C & D \end{bmatrix}$. Then
\[
A_i = HA_{i-1}.
\]
Thus
\[
A_i = H^i A_0, \quad i \geq 0
\]
where \( A_0 = \begin{bmatrix} \pi_1 \\ \pi_i' \\ \pi_0 \end{bmatrix} = \begin{bmatrix} \lambda_1 - \frac{\mu_1 \lambda_2}{\mu_2 (\lambda_1 + \mu_2)} \\ \frac{\mu_1 \lambda_2}{\lambda_1 + \mu_2} \\ 1 \end{bmatrix} \pi_0 \), using the first two equations of system (24).

(25) shows that in order to find the stability criterion of the system in (24) we first need to study the properties of \( H \).

For that we compute its eigenvalues \( l_0, l_1, l_2, l_3 \) by solving the characteristic equation \( |I_4 - H| = 0 \). This leads to

\[
|I_4 - H| = l(l - 1)(l^2 - l(a_1 + a_5 - 1) + a_3 a_5). \tag{26}
\]

\( H \) has two obvious eigenvalues \( l_0 = 0 \) and \( l_3 = 1 \). To find the last two eigenvalues, let’s find the root of the quadratic polynomial

\[
p(l) = l^2 - l(a_1 + a_5 - 1) + a_3 a_5. \tag{27}
\]

It can be shown through simple algebra that the discriminant of the above polynomial is strictly positive. Hence the remaining eigenvalues \( l_1 \) and \( l_2 \) are real and distinct. Let’s assume that \( l_1 < l_2 \). This means that the matrix \( H \) is diagonalizable and can be written as

\[
H = BAB^{-1},
\]

where the columns of \( B \) are the eigenvectors of \( H \) and form a basis of \( \mathbb{R}^4 \). We denote by \( e_0, e_1, e_2, e_3 \) the eigenvectors corresponding to \( l_0, l_1, l_2, l_3 \).

So we can write \( A_0 \) as

\[
A_0 = (a_0 e_0 + a_1 e_1 + a_2 e_2 + a_3 e_3) \pi_0, \tag{28}
\]

with \( a_0, a_1, a_2, a_3 \in \mathbb{R} \). Hence for \( i > 0 \),

\[
A_i = H^i A_0 = (a_0 H^i e_0 + a_1 H^i e_1 + a_2 H^i e_2 + a_3 H^i e_3) \pi_0 = (a_0 l_1^i e_0 + a_1 l_1^i e_1 + a_2 l_1^i e_2 + a_3 l_1^i e_3) \pi_0 = (a_1 l_1^i e_1 + a_2 l_1^i e_2 + a_3 e_3) \pi_0,
\]

since \( l_0 = 0 \) and \( l_3 = 1 \). Equation (29) shows that three conditions need to be satisfied for the system to be stable and a steady-state distribution to exist:

- **Condition 1**: \( |l_1| < 1 \) and \( |l_2| < 1 \).
- **Condition 2**: \( a_3 = 0 \).
- **Condition 3**: \( a_1 l_1^i e_1 + a_2 l_1^i e_2 \) has positive components for all \( i > 0 \).

**Condition 1** and **Condition 2** ensure that

\[
\lim_{i \to \infty} \pi_i = \lim_{i \to \infty} \pi_i' = 0
\]

and thus the sum of all probabilities, \( \pi_0 + \sum_{i=1}^{\infty} (\pi_i + \pi_i') \), does not diverge. **Condition 3** makes sure that the components of \( A_0 \) are positive probabilities. We will show that (23) is sufficient for the above three conditions to hold.

Given that \( l_1 \) and \( l_2 \) are the roots of (27), then the following holds

\[
l_1 l_2 = a_3 a_5 \quad l_1 + l_2 = a_1 + a_5 - 1. \tag{30}
\]

However, \( l_1 l_2 = a_3 a_5 = \frac{\lambda_1^2}{\mu_1 (\mu_2 + \lambda_1)} \geq 0 \). This means that either both \( l_1 \) and \( l_2 \) are positive or they are both negative. Using (30) again, we notice that

\[
l_1 + l_2 = a_1 + a_5 - 1 = \frac{\lambda_1 \mu_2 + \lambda_1^2 + \lambda_1 \lambda_2 + \lambda_1 \mu_1}{\mu_1 (\mu_2 + \lambda_1)} \geq 0.
\]

This shows that both \( l_1 \) and \( l_2 \) are strictly positive (since 0 is not a root of \( p(l) \)). So to prove that **Condition 1** is satisfied we need to prove that \( l_1 < l_2 < 1 \). This is equivalent to show that (27) evaluated at 1 is strictly positive and that \( l_1 l_2 < 1 \) since \( p(l) \) is a convex quadratic function in \( l > 0 \). Using simple algebra it can be shown that

\[
p(1) = 1 - (a_1 + a_5 - 1) + a_3 a_5 = \frac{\mu_1 \mu_2 - \lambda_1 (2 a_1 + a_5 - 1)}{\mu_1 (\mu_2 + \lambda_1)} > 0,
\]

where the last inequality is due to (23). Moreover, (23) tells us that \( \mu_1 \) should be strictly bigger that \( \lambda_1 \). Thus we get that

\[
l_1 l_2 = \frac{\lambda_1}{\mu_1 \mu_2 + \lambda_1} < 1.
\]

This shows that \( 0 < l_1 < l_2 < 1 \) and that **Condition 1** is satisfied.

To prove **Condition 2** we start by computing the eigenvectors of \( H \). For \( l_0 = 0 \), we solve the system given by \( H e_0 = 0 \). If \( e_0 = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \) then

\[
\begin{bmatrix}
 a_1 & -a_2 & -a_3 & 0 \\
 a_4 & a_5 & 0 & 0 \\
 1 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0
\end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.
\]

This system leads to \( e_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}^T \). Similarly, for \( j = 1, 2, 3 \), if \( e_j = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ 1 \end{bmatrix}^T \) then solving the system

\[
\begin{bmatrix}
 a_1 & -a_2 & -a_3 & 0 \\
 a_4 & a_5 & 0 & 0 \\
 1 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0
\end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 \\ u_2 \\ u_3 \end{bmatrix}
\]

leads to \( e_j = \begin{bmatrix} l_j (l_j - a_5) \\ l_j a_4 \\ l_j - a_5 \\ a_4 \end{bmatrix}^T \).

We know that \( H = BAB^{-1} \). If

\[
A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & l_2 & 0 & 0 \\ 0 & 0 & l_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},
\]

then

\[
B = \begin{bmatrix}
 1 - a_5 & l_2 (l_2 - a_5) & l_1 (l_1 - a_5) & 0 \\
 a_4 & l_2 a_4 & l_1 a_4 & 0 \\
 1 - a_5 & l_2 - a_5 & l_1 - a_5 & 0 \\
 a_4 & a_4 & a_4 & 1
\end{bmatrix}.
\]

Note that the determinant of \( B \), \( |B| \), is non-zero when we assume (23). Indeed,

\[
|B| = a_4 a_5 (l_2 - l_1) (-2 + a_5 - a_3 a_5 + a_1) < 0
\]
since \( l_2 > l_1 \) and \(-2 + a_5 - a_3 a_5 + a_1 = -p(1) < 0\) as shown before. In order to compute \( \alpha_3 \), we rewrite (28) as follows

\[
\mathbf{A}_0 = \begin{bmatrix} e_3 & e_2 & e_1 & e_0 \end{bmatrix} \begin{bmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} \pi_0 = B \begin{bmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} \pi_0.
\]

But we also know that

\[
\mathbf{A}_0 = \begin{bmatrix} \frac{\lambda}{\mu_1} - \frac{\mu_2 \lambda_2}{\lambda_1 + \mu_2} \\ \frac{\lambda_1 + \mu_2}{1} \\ 0 \\ 0 \end{bmatrix} \pi_0 = \begin{bmatrix} a_1 - 1 \\ a_4 \\ 1 \\ 0 \end{bmatrix} \pi_0.
\]

Thus

\[
\mathbf{B} \begin{bmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} a_1 - 1 \\ a_4 \\ 1 \\ 0 \end{bmatrix}.
\]

(31) Solving the system in (31) with respect to \( \alpha_3, \alpha_2, \alpha_1 \) and \( \alpha_0 \) we get that

\[
\begin{bmatrix} \alpha_3 \\ \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{l_2 - l_1} \\ 0 \\ 0 \end{bmatrix} \pi_0.
\]

Thus \( \alpha_3 = 0 \) and Condition 2 is proved. Note that we didn’t need any assumptions to prove this condition.

Given the above results, we can now rewrite the system in (29) as

\[
\begin{align*}
\mathbf{A}_i &= (\alpha_2 l_2^i e_2 + \alpha_1 l_1^i e_1) \pi_0, \quad i > 0 \\
\mathbf{A}_0 &= (\alpha_2 e_2 + \alpha_1 e_1) \pi_0 = \begin{bmatrix} a_1 - 1 \\ a_4 \\ 1 \\ 0 \end{bmatrix} \pi_0.
\end{align*}
\]

(32)

Using (32) we can prove Condition 3. In fact, for any \( i > 0 \),

\[
\begin{align*}
\alpha_2 l_2^i e_2 + \alpha_1 l_1^i e_1 &= (a) \alpha_2 \left( l_2^i e_2 - l_1^i e_1 \right) \\
&\geq (b) \alpha_2 l_2^i (e_2 - e_1) \\
&\overset{(c)}{=} l_1^i \begin{bmatrix} a_1 - 1 \\ a_4 \\ 1 \\ 0 \end{bmatrix} \\
&\overset{(d)}{=} 0,
\end{align*}
\]

where \( x \succ y \) for some vectors \( x \) and \( y \) means that the components of \( x - y \) are strictly positive and

\begin{itemize}
  \item \( (a) \) is because \( \alpha_2 = -\alpha_1 \),
  \item \( (b) \) is because \( 0 < l_1 < l_2 \),
  \item \( (c) \) is obtained from the second equality in (32),
  \item \( (d) \) follows since \( a_1 - 1 > 0 \) and \( a_4 > 0 \).
\end{itemize}

Up till now we have shown that if \( \mu_1 > \lambda_1 \left( 1 + \frac{d_2}{\mu_4} \right) \), the system described in Section II is stable and a steady-state distribution exists given by (32). The final point to prove in Theorem [1] is the expression of \( \pi_0 \). For that we solve for \( \pi_0 \) the following equation

\[
\pi_0 + \sum_{i=1}^{\infty} \pi_i + \pi_0' = \pi_0 + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \sum_{i=1}^{\infty} \mathbf{A}_i = 1.
\]

Using the first equation of (32) and replacing \( \alpha_1 \) and \( \alpha_2 \) by their expressions in function of \( l_1 \) and \( l_2 \), using the fact that \( l_1 + l_2 \) and \( l_1 l_2 \) are given by (30) and finally replacing \( a_1, a_2, a_3, a_4 \) and \( a_5 \) by their expressions in function of \( \lambda_1, \lambda_2, \mu_1, \mu_2 \) we get

\[
\pi_0 = \frac{\mu_2}{\mu_2 + \lambda_2} - \frac{\lambda_1}{\mu_1}.
\]

\[\square\]

Proof. (Corollary [1]) At any point in time, there are exactly \( i \) stream \( U_i \) packets in the system if we are in state \( q_i \) or \( q_{i+1} \) in the Markov chain given by Fig. 2. This means that the probability of having exactly \( i \) stream \( U_i \) packets in the system is \( \pi_i + \pi_i' + 1 \). Hence, using the same quantities as in the previous proof

\[
\phi(\epsilon) = E(e^{s N(t)}) = \sum_{n=0}^{\infty} e^{sn} (\pi_i + \pi_i') = \sum_{n=0}^{\infty} e^{sn} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \sum_{n=0}^{\infty} e^{sn} e^{s \lambda_1 t} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \alpha_2 \pi_0 \begin{bmatrix} \sum_{n=0}^{\infty} (e^{s \lambda_1 t})^{n} e_2^T \\ 1 \\ 0 \\ 0 \end{bmatrix} = \alpha_2 \pi_0 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \alpha_2 \pi_0 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \alpha_2 \pi_0 \begin{bmatrix} \frac{a_4 + 1 - a_5 e^{s \lambda_1 t}}{1 - l_2 e^{s \lambda_1 t}} \end{bmatrix}.
\]

(33)

where the quantities used here are the one defined in the proof of Theorem [1] Thus,

\[
\phi(\epsilon) = \pi_0 \begin{bmatrix} \mu_1 \lambda_2 + \mu_2 - \lambda_1 e^{s \lambda_1 t} \\ \mu_1 \mu_2 - \lambda_2 e^{s \lambda_1 t} - \lambda_1 e^{s \lambda_2 t} - \lambda_2 e^{s \lambda_2 t} \\ \mu_1 \mu_2 - \lambda_2 e^{s \lambda_1 t} - \lambda_1 e^{s \lambda_2 t} - \lambda_2 e^{s \lambda_2 t} \\ 0 \end{bmatrix}.
\]

This last equality is obtained by using (30), \( \alpha_2 = \frac{1}{l_2 - l_1} \) and replacing \( a_1, a_2, a_3, a_4, a_5 \) by their expressions in function of \( \lambda_1, \lambda_2, \mu_1 \) and \( \mu_2 \) in (33). Finally,

\[
\mathbb{E}(N(t)) = \frac{d\phi(\epsilon)(s)}{ds} \bigg|_{s=0} = \lambda_1 \left( 2 \lambda_2 \mu_2 + \lambda_2 \mu_1 + \frac{\lambda_2^2 + \mu_2^2}{\mu_2 + \lambda_2} \right) - \lambda_1 \left( \mu_1 \mu_2 - \lambda_1 \left( \mu_2 + \lambda_2 \right) \right).
\]

\[\square\]