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' ([1]) 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

$$\Delta = \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau \Delta(t)\mathrm{d}t, \tag{1}$$

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 Δ were investigated under the assumption that the packets are generated by a Poisson process, and various transmission policies (M/M/1, M/M/∞, 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

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

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 blocking 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 blocking policy, if the generated update finds the system busy, it is discarded. 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 λ, 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 μ_1, a packet of the priority scheme at rate μ_2.

We will study the above model and answer the questions: what should the relation between λ_1, μ_1, λ_2 and μ_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 λ. 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 $\mathcal{U}_1$ and $\mathcal{U}_2$ 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 μ_1 for stream $\mathcal{U}_1$ and rate μ_2 for stream $\mathcal{U}_2$. 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 $\mathcal{U}_1$ should be sent. Hence the server applies a FCFS policy on the packets from stream $\mathcal{U}_1$ with a buffer to save waiting updates. On the other hand, we assume that the information carried by stream $\mathcal{U}_2$ 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 $\mathcal{U}_2$ is generated. This means that if a newly generated packet from stream $\mathcal{U}_2$ finds the system busy (serving a packet from $\mathcal{U}_1$ or $\mathcal{U}_2$), the server preempts the update currently in service and start serving the new packet. Moreover, if the preempted packet belongs to $\mathcal{U}_1$, this packet is placed back at the head of the $\mathcal{U}_1$-buffer so that it can be served once the system is idle again. If the preempted packet belongs to $\mathcal{U}_2$ then it is discarded. However, if a newly generated $\mathcal{U}_1$-packet finds the system busy serving a $\mathcal{U}_2$-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 $\mathcal{U}_1$. In this plot, t_i and D_i refer to the generation and delivery times of the i^{th} packet of stream $\mathcal{U}_1$ 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 $\mathcal{U}_2$ only.

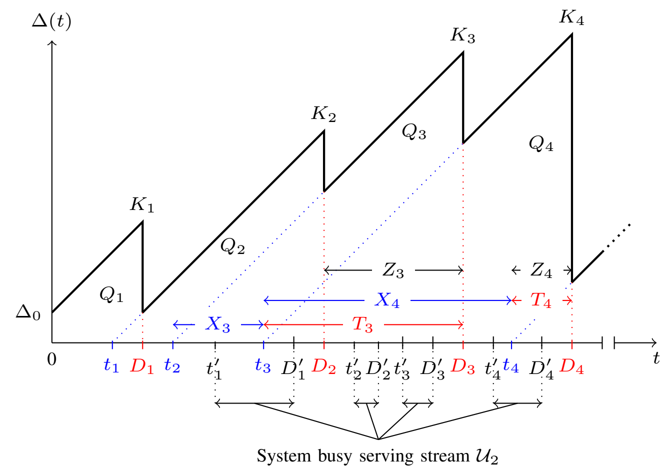


Fig. 1. Variation of the instantaneous age of stream $\mathcal{U}_1$.

III. System Stability and Stationary Distribution

The fact that we wish to receive all of stream $\mathcal{U}_1$ updates and that stream $\mathcal{U}_2$ has higher priority and preempts stream $\mathcal{U}_1$ 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 $\mathcal{U}_1$ while states q'_i, $i > 0$, in the row below correspond to the queue serving a packet from stream $\mathcal{U}_2$. In both cases there are $i - 1$ stream $\mathcal{U}_1$ updates waiting in the buffer.

The system leaves state q_0 at rate λ_1 to state q_1 when a packet from stream $\mathcal{U}_1$ is generated first and it leaves q_0 at rate λ_2 to state q'_1 when a packet from stream $\mathcal{U}_2$ is generated first. However, when the system enters state q_i, $i > 0$, three exponential clocks start: a clock with rate μ_1 which corresponds to the service time of the stream $\mathcal{U}_1$ packet being served, a clock with rate λ_1 which corresponds to the generation time of stream $\mathcal{U}_1$ packets and a clock with rate λ_2 which corresponds to the generation time of stream $\mathcal{U}_2$ packets. If the μ_1-clock ticks first, the system goes to state q_{i-1}: this means that the current stream $\mathcal{U}_1$ packet was delivered and the queue begins the service of the next one in the buffer (if there is any). However, if the λ_1-clock ticks first, a new stream $\mathcal{U}_1$ update is generated and added to the buffer and hence the system goes to state q_{i+1}. On the other hand, if the λ_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 $\mathcal{U}_2$ 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 λ_1 and a clock with rate μ_2 which corresponds to the

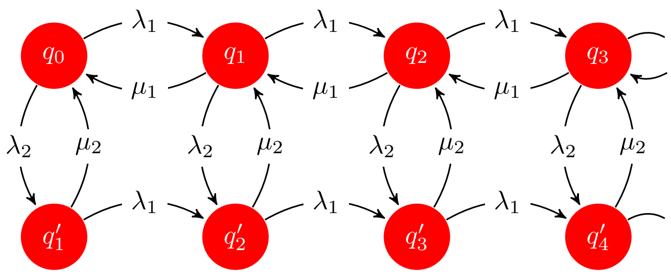


Fig. 2. Markov chain governing the number of packets in the system.

service time of a stream $\mathcal{U}_2$ packet. If the λ_1-clock ticks first, the newly generated stream $\mathcal{U}_1$ packet is placed in the buffer and the stream $\mathcal{U}_2$ update is continued to be served. Hence the system goes to state q'_{i+1}. However, if the μ_2-clock ticks first, the stream $\mathcal{U}_2$ packet has finished service and the system starts serving the first stream $\mathcal{U}_1$ 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*

$$\mu_1 > \lambda_1\left(1+\frac{\lambda_2}{\mu_2}\right). \tag{3}$$

In this case the Markov chain shown in Fig. 2 has a stationary distribution $\Pi = [\pi_0, \pi_1, \dots, \pi_i, \dots, \pi'_1, \dots, \pi'_i, \dots]$, *where* π_i *denotes the stationary probability of state* q_i, $i \geq 0$, *and* π'_i *denotes the stationary probability of state* q'_i, $i > 0$. *This stationary distribution is described by the following system of equations,*

$$\pi_0 = \frac{\mu_2}{\mu_2+\lambda_2} - \frac{\lambda_1}{\mu_1}, \tag{4}$$

$$A_i = \begin{bmatrix} 0 & I \end{bmatrix} H^i \begin{bmatrix} \frac{\lambda}{\mu_1} - \frac{\mu_2\lambda_2}{\mu_1(\lambda_1+\mu_2)} \\ \frac{\lambda_2}{\lambda_1+\mu_2} \\ 1 \\ 0 \end{bmatrix} \pi_0, \ \ i \geq 1 \tag{5}$$

where $\lambda = \lambda_1 + \lambda_2$,

$$A_i = \begin{bmatrix} \pi_i \\ \pi'_i \end{bmatrix}, \ H = \begin{bmatrix} C & D \\ I & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 + \frac{\lambda}{\mu_1} - \frac{\mu_2\lambda_2}{\mu_1(\mu_2+\lambda_1)} & -\frac{\mu_2\lambda_1}{\mu_1(\mu_2+\lambda_1)} \\ \frac{\lambda_2}{\mu_2+\lambda_1} & \frac{\lambda_1}{\mu_2+\lambda_1} \end{bmatrix}, \ D = \begin{bmatrix} -\frac{\lambda_1}{\mu_1} & 0 \\ 0 & 0 \end{bmatrix},$$

I is the 2×2 identity matrix and 0 is the 2×2 zero matrix.

Proof. The distribution given by (4) and (5) satisfy the detailed balance equations of the Markov chain shown in Fig. 2. Moreover, (3) is the condition needed to have $\pi_0 > 0$. □

The condition in (3) can be interpreted as follows: define the map f from the state-space of the chain as $f(s) = 0$ if s is in $\{q_0, q_1, \dots\}$ and $f(s) = 1$ if $s \in \{q'_1, q'_2, \dots\}$. For each s and s' for which $f(s) = 0$ and $f(s') = 1$ the transition rate from s to s' is the same (λ_2) and similarly for s and s' with $f(s) = 1$, $f(s') = 0$, (μ_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 λ_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 (3).

IV. Ages of streams $\mathcal{U}_1$ and $\mathcal{U}_2$

A. Preliminaries

In this section, unless stated otherwise, all random variables correspond to stream $\mathcal{U}_1$. 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 $\mathcal{U}_1$. To differentiate between streams we will use superscripts, so $U^{(i)}$ corresponds to the steady-state variable U relative to stream $\mathcal{U}_i$, $i = 1, 2$.

In addition to that, we define: (i) $X^{(i)}$ to be the interarrival time between two consecutive generated updates from stream $\mathcal{U}_i$, 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 $\mathcal{U}_i$ updates, so $f_{S^{(i)}}(t) = \mu_i e^{-\mu_i t}$, $i = 1, 2$ (iii) T_j to be the system time, or the time spent by the j^{th} stream $\mathcal{U}_1$ 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 see that from the point of view of stream $\mathcal{U}_1$ the system behaves as an M/G/1 queue, with each packet j having an independent "virtual" service time Z_j which could be different from its "physical" service time $S_j^{(1)}$. We define the "virtual" service time Z_j as follows:

$$Z_j = D_j - \max(D_{j-1}, t_j), \tag{6}$$

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.

B. Average peak age of stream $\mathcal{U}_1$

For stream $\mathcal{U}_1$, 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. We start by defining 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.

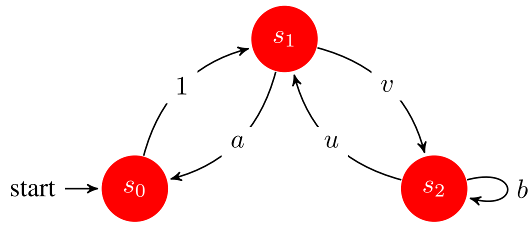


Fig. 3. Semi-Markov chain representing the "virtual" service time Y_j.

Lemma 1. *Let Y_j be the "virtual" service time of packet j given that this packet does not finds the system in state q_1', i.e. $\mathbb{P}(Y_j > t) = \mathbb{P}\left(Z_j > t|\overline{\Psi_j}\right)$. Then, in steady state,*

$$\phi_Y(s) = \mathbb{E}\left(e^{sY}\right) = \frac{\mu_1(\mu_2 - s)}{s^2 - s(\mu_2 + \mu_1 + \lambda_2) + \mu_1\mu_2}. \quad (7)$$

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. $\mathbb{P}\left(Y_j' > t\right) = \mathbb{P}(Z_j > t|\Psi_j)$. Then, in steady state,

$$\phi_{Y'}(s) = \mathbb{E}\left(e^{sY'}\right) = \frac{\mu_1\mu_2}{s^2 - s(\mu_2 + \mu_1 + \lambda_2) + \mu_1\mu_2}. \quad (8)$$

Proof. We start by proving (7). 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 $\mathcal{U}_1$. 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 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 $a = \mathbb{P}\left(S^{(1)} < X^{(2)}\right)$ and its value A has distribution $\mathbb{P}(A > t) = \mathbb{P}\left(S^{(1)} > t|S^{(1)} < X^{(2)}\right)$. At this point the stream $\mathcal{U}_1$ 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 $v = 1 - a = \mathbb{P}\left(X^{(2)} < S^{(1)}\right)$ and its value V has distribution $\mathbb{P}(V > t) = \mathbb{P}\left(X^{(2)} > t|X^{(2)} < S^{(1)}\right)$. At this point, a new stream $\mathcal{U}_2$ update is generated and preempts the stream $\mathcal{U}_1$ packet currently in service. In this case the system goes to state s_2, where the preempted stream $\mathcal{U}_1$ update is placed back at the head of the buffer and the system starts service of the stream $\mathcal{U}_2$ update.

When the system arrives in state s_2, this means a new stream $\mathcal{U}_2$ packet was just generated and is starting service. Thus, two clocks start: a service clock $S^{(2)}$ and a clock $X^{(2)}$. The service clock ticks first with probability $u = \mathbb{P}\left(S^{(2)} < X^{(2)}\right)$ and its value U has distribution $\mathbb{P}(U > t) = \mathbb{P}\left(S^{(2)} > t|S^{(2)} < X^{(2)}\right)$. At this point, the packet currently being served finishes service before any new stream $\mathcal{U}_2$ packet is generated and the system goes back to state s_1 where the j^{th} packet of stream $\mathcal{U}_1$ starts service again. However, clock $X^{(2)}$ ticks first with probability $b = 1 - u$ and its value $B =$ has distribution $\mathbb{P}(B > t) = \mathbb{P}\left(X^{(2)} > t|X^{(2)} < S^{(2)}\right)$. At this point, a new stream $\mathcal{U}_2$ 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_j, B_j, U_j, V_j 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 $\{A_k, B_k, U_k, V_k\}$ 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,

$$\begin{aligned}\phi_Y(s) &= \mathbb{E}\left(\mathbb{E}\left(e^{sY}|\,(I_1, I_2, I_3, I_4) = (i_1, i_2, i_3, i_4)\right)\right) \\ &= \sum_{i_1,i_2,i_3,i_4} \Big[a^{i_1}b^{i_2}u^{i_3}v^{i_4}Q(i_1, i_2, i_3, i_4) \\ &\quad \mathbb{E}\left(e^{s\left(\sum_{k=1}^{i_1} A_k + \sum_{k=1}^{i_2} B_k + \sum_{k=1}^{i_3} U_k + \sum_{k=1}^{i_4} V_k\right)}\right)\Big] \\ &= \sum_{i_1,i_2,i_3,i_4} \Big[a^{i_1}b^{i_2}u^{i_3}v^{i_4}Q(i_1, i_2, i_3, i_4) \\ &\quad \mathbb{E}\left(e^{sA}\right)^{i_1}\mathbb{E}\left(e^{sB}\right)^{i_2}\mathbb{E}\left(e^{sU}\right)^{i_3}\mathbb{E}\left(e^{sV}\right)^{i_4}\Big]. \end{aligned} \quad (9)$$

However (9) is nothing but the generating function $H_1(D_1, D_2, D_3, D_4)$ of the detour flow graph shown in Fig. 4a, where D_1, D_2, D_3, D_4 are dummy variables (see [11, pp. 213–216]). Simple calculations give

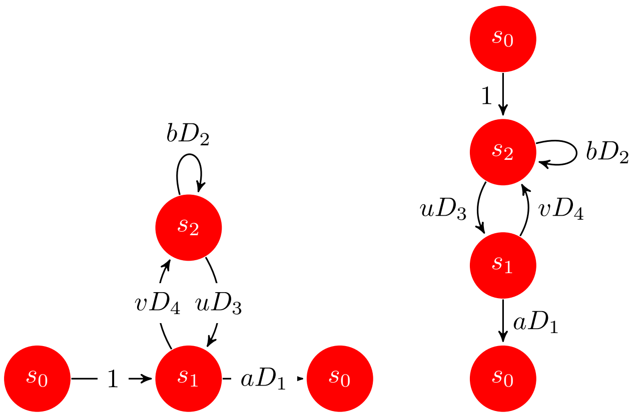
$$\begin{aligned} &H_1(D_1, D_2, D_3, D_4) \\ &= \sum_{i_1,i_2,i_3,i_4} \left[Q(i_1, i_2, i_3, i_4)a^{i_1}b^{i_2}u^{i_3}v^{i_4}D_1^{i_1}D_2^{i_2}D_3^{i_3}D_4^{i_4}\right] \\ &= \frac{aD_1(1 - bD_2)}{1 - bD_2 - uD_3vD_4}. \end{aligned} \quad (10)$$

Thus

$$\phi_Y(s) = H_1\left(\mathbb{E}\left(e^{sA}\right), \mathbb{E}\left(e^{sB}\right), \mathbb{E}\left(e^{sU}\right), \mathbb{E}\left(e^{sV}\right)\right).$$

From [7, Appendix A, Lemma 2], we know that A, B, U and V are exponentially distributed with $\mathbb{E}\left(e^{sB}\right) = \mathbb{E}\left(e^{sU}\right) = \frac{\lambda_2+\mu_2}{\lambda_2+\mu_2-s}$ and $\mathbb{E}\left(e^{sA}\right) = \mathbb{E}\left(e^{sV}\right) = \frac{\lambda_2+\mu_1}{\lambda_2+\mu_1-s}$. Simple computations show that $a = \frac{\mu_1}{\mu_1+\lambda_2}$, $b = \frac{\lambda_2}{\mu_2+\lambda_2}$, $u = \frac{\mu_2}{\mu_2+\lambda_2}$, $v = \frac{\lambda_2}{\mu_1+\lambda_2}$. Finally, replacing the above expressions into (10), we get our result.

To prove (8), we use the same method as before but in this case we notice that the j^{th} packet from stream $\mathcal{U}_1$ finds the



(a) (b)

Fig. 4. Detour flow graphs for (a) Y and (b) Y'.

system busy serving a packet from stream $\mathcal{U}_2$. This translates in the detour flow graph shown in Fig. 4b. The generating function of this graph is

$$H_2(D_1, D_2, D_3, D_4) = \frac{aD_1uD_3}{1 - bD_2 - vD_4uD_3}. \tag{11}$$

For $(D_1, D_2, D_3, D_4) = \left(\mathbb{E}\left(e^{sA}\right), \mathbb{E}\left(e^{sB}\right), \mathbb{E}\left(e^{sU}\right), \mathbb{E}\left(e^{sV}\right)\right)$ and replacing a, b, u and v by their values in (11), we get (8). □

Lemma 2. *The first and second moments of the "virtual" service time Z are given by*

$$\mathbb{E}(Z) = \frac{\lambda_2}{(\lambda_1+\mu_2)(\mu_2+\lambda_2)} + \frac{\lambda_1+\lambda_2+\mu_2}{\mu_1(\lambda_1+\mu_2)},$$
$$\mathbb{E}\left(Z^2\right) = \frac{2\left((\lambda_2+\mu_2)^2(\lambda_2+\mu_2+\lambda_1) + \lambda_2\mu_1(2\lambda_2+\mu_1+2\mu_2)\right)}{\mu_1^2\mu_2(\lambda_1+\mu_2)(\lambda_2+\mu_2)} \tag{12}$$

Proof. For any packet j of stream $\mathcal{U}_1$, conditioning on the event Ψ_j, we get

$$\begin{aligned}\mathbb{E}(Z_j) &= \mathbb{P}(\Psi_j)\,\mathbb{E}(Z_j|\Psi_j) + \mathbb{P}\left(\overline{\Psi_j}\right)\mathbb{E}\left(Z_j|\overline{\Psi_j}\right)\\ &= \mathbb{P}(\Psi_j)\,\mathbb{E}\left(Y_j'\right) + \mathbb{P}\left(\overline{\Psi_j}\right)\mathbb{E}(Y_j),\end{aligned} \tag{13}$$

where Y_j' and Y_j are defined as in Lemma 1. From Theorem 1 we deduce that $\mathbb{P}(\Psi_j) = \pi_1' = \frac{\lambda_2}{\lambda_1+\mu_2}\pi_0$. So in steady-state, $\mathbb{E}(Z) = \pi_1'\mathbb{E}(Y') + (1-\pi_1')\mathbb{E}(Y)$. Moreover, using (7) and (8) we get

$$\mathbb{E}(Y) = \frac{\mu_2+\lambda_2}{\mu_1\mu_2},\ \mathbb{E}(Y') = \frac{\mu_1+\mu_2+\lambda_2}{\mu_1\mu_2}.$$

Similarly, $\mathbb{E}\left(Z^2\right) = \pi_1'\mathbb{E}\left(Y'^2\right) + (1-\pi_1')\mathbb{E}\left(Y^2\right)$. Using (7) and (8) we get

$$\mathbb{E}\left(Y^2\right) = \frac{2\left((\mu_2+\lambda_2)^2 + \mu_1\lambda_2\right)}{(\mu_1\mu_2)^2} \text{ and}$$

$$\mathbb{E}\left(Y'^2\right) = \frac{2\left((\mu_1+\mu_2+\lambda_2)^2 - \mu_1\mu_2\right)}{(\mu_1\mu_2)^2}.$$

□

It is worth noting that the condition given in (3) leads to the inequality $\lambda_1\mathbb{E}(Z) < 1$, which implies the stability of the M/G/1 queue.

Theorem 2. *The average peak age of stream $\mathcal{U}_1$ is given by*

$$\Delta_{peak,1} = \frac{1}{\lambda_1} + \mathbb{E}(Z) + \frac{\lambda_1\mathbb{E}\left(Z^2\right)}{2\left(1-\lambda_1\mathbb{E}(Z)\right)}, \tag{14}$$

where Z is the steady state equivalent of Z_j defined in (6), and λ_1 is the generation rate of updates belonging to $\mathcal{U}_1$.

With the expressions for $\mathbb{E}(Z)$ and $\mathbb{E}\left(Z^2\right)$ given in Lemma 2, we thus obtain the average peak age of stream $\mathcal{U}_1$ in closed form.

Proof. As we have seen before, the system from stream $\mathcal{U}_1$ point of view acts like an M/G/1 queue with service time Z. Applying [8, Proposition 2] for a single stream M/G/1 system with update rate λ_1 and service time Z gives (14) as the average peak age for stream $\mathcal{U}_1$. □

C. Lower bound on the average age of stream $\mathcal{U}_1$

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

From (8), we can deduce that $Y' = X^{(2)} + Y$ with $X^{(2)}$ and Y being independent. In fact,

$$\mathbb{E}\left(e^{sY'}\right) = \mathbb{E}\left(e^{sX^{(2)}}\right)\mathbb{E}\left(e^{sY}\right),$$

where $\mathbb{E}\left(e^{sX^{(2)}}\right) = \frac{\mu_2}{\mu_2-s}$ and $E\left(e^{sY}\right)$ given by (7). This explains why $\mathbb{E}(Y') = \mathbb{E}(Y) + \frac{1}{\mu_2} > \mathbb{E}(Y)$. Since $\mathbb{E}(Z) = \pi_1'\mathbb{E}(Y') + (1-\pi_1')\mathbb{E}(Y)$, then

$$\mathbb{E}(Y) < \mathbb{E}(Z) < \mathbb{E}(Y').$$

Hence the average age of an M/G/1 queue with service time Y will give us a lower bound to the average age of our M/G/1 system with service time Z.

Lemma 3. *Assume an M/G/1 queue with interarrival time $X^{(1)}$ exponentially distributed with rate λ_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*

$$f_T(t) = -C_1e^{-\alpha_1 t} - C_2e^{-\alpha_2 t},\ t \geq 0, \tag{15}$$

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

$$s^2 - s(\mu_1+\mu_2+\lambda_2-\lambda_1) + \mu_1\mu_2 - \lambda_1\mu_2 - \lambda_1\lambda_2,$$

$$C_1 = \frac{(1-\rho)\mu_1(\mu_2-\alpha_1)}{\alpha_1-\alpha_2},\ C_2 = \frac{(1-\rho)\mu_1(\mu_2-\alpha_2)}{\alpha_2-\alpha_1}$$

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

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 (7) we get

$$\phi_T(s) = \frac{(1-\rho)\mu_1(\mu_2 - s)}{s^2 - s(\mu_1+\mu_2+\lambda_2-\lambda_1) + \mu_1\mu_2 - \lambda_1\mu_2 - \lambda_1\lambda_2}$$
$$= \frac{C_1}{s-\alpha_1} + \frac{C_2}{s-\alpha_2}, \tag{16}$$

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 α_1 and α_2 are positive. Without loss of generality, we take $\alpha_1 > \alpha_2$. Taking the inverse Laplace transform of $\phi_T(-s)$ we get (15). $\square$

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

$$\Delta_{LB} = \lambda_1 \left(\frac{1}{2}\mathbb{E}\left({X_j^{(1)}}^2\right) + \mathbb{E}\left(T_j X_j^{(1)}\right)\right), \tag{17}$$

where for the j^{th} packet we have $T_j = (T_{j-1} - X_j^{(1)})^+ + Y_j$, $f(x) = (x)^+ = x1_{\{x\geq 0\}}$ and $1_{\{.\}}$ is the indicator function. So $\mathbb{E}\left(T_j X_j^{(1)}\right)$ becomes

$$\mathbb{E}\left(T_j X_j^{(1)}\right) = \mathbb{E}\left(X_j^{(1)}(T_{j-1} - X_j^{(1)})^+\right) + \mathbb{E}\left(Y_j\right)\mathbb{E}\left(X_j^{(1)}\right), \tag{18}$$

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

Proposition 1.

$$\mathbb{E}\left(X_j^{(1)}(T_{j-1} - X_j^{(1)})^+\right)$$
$$= \frac{\lambda_1(1-\rho)\mu_1\left(\left((\alpha_1+\alpha_2)^2 - \alpha_1\alpha_2\right)(\mu_2 + 2\lambda_1\mu_2 - \alpha_1\alpha_2)\right)}{\mu_1^2(\lambda_1+\mu_2)^2(\alpha_1\alpha_2)^2}$$
$$+ \frac{\lambda_1(1-\rho)\mu_1\left((\alpha_1+\alpha_2)\left(\lambda_1^2\mu_2 - 2\lambda_1\alpha_1\alpha_2\right) - \lambda_1^2\alpha_1\alpha_2\right)}{\mu_1^2(\lambda_1+\mu_2)^2(\alpha_1\alpha_2)^2}, \tag{19}$$

where $\alpha_1+\alpha_2 = \mu_1+\mu_2+\lambda_2-\lambda_1$, $\alpha_1\alpha_2 = \mu_1\mu_2 - \lambda_1\mu_2 - \lambda_1\lambda_2$ *and* $\rho = \lambda_1\mathbb{E}\left(Y\right) = \frac{\lambda_1(\mu_2+\lambda_2)}{\mu_1\mu_2}$.

Proof. Given that T_{j-1} and $X_j^{(1)}$ are independent then

$$\mathbb{E}\left(X_j^{(1)}(T_{j-1} - X_j^{(1)})^+\right)$$
$$= \int_0^\infty \int_x^\infty x(t-x) f_T(t) \lambda_1 e^{-\lambda_1 x} \mathrm{d}t\mathrm{d}x$$

Replacing $f_T(t)$ by its value in (15), we get (19) after some computations. $\square$

Finally, using (19), $\mathbb{E}\left(Y_j\right) = \mathbb{E}\left(Y\right) = \frac{\mu_2+\lambda_2}{\mu_1\mu_2}$ and $\mathbb{E}\left(X_j^{(1)}\right) = \mathbb{E}\left(X^{(1)}\right) = \frac{1}{\lambda_1}$, we can find a closed form expression for $\mathbb{E}\left(T_j X_j^{(1)}\right)$. Replacing this expression in (17) and using the fact that $\mathbb{E}\left({X_j^{(1)}}^2\right) = \frac{2}{\lambda_1^2}$, we obtain a closed form expression of the average age Δ_{LB} of an M/G/1 queue with interarrival time $X^{(1)}$ and service time Y. This is also a lower bound on the average age of the M/G/1 queue with interarrival time $X^{(1)}$ and service time Z.

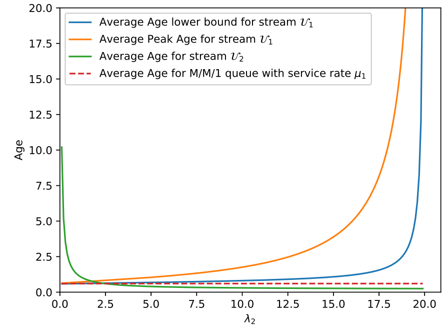


Fig. 5. Plot of the average age for stream $\mathcal{U}_2$ and average peak age and lower bound on the average age for stream $\mathcal{U}_1$, with $\mu_1 = 10$, $\mu_2 = 5$, $\lambda_1 = 2$ and $\lambda_2 < \frac{\mu_1\mu_2}{\lambda_1} - \mu_2 = 20$.

D. Average age of stream $\mathcal{U}_2$

By design, stream $\mathcal{U}_2$ is not interfered at all by stream $\mathcal{U}_1$ and hence behaves like a traditional M/M/1/1 with preemption queue with generation rate λ_2 and service rate μ_2. The average age of this stream was computed in [2] to be

$$\Delta_{\mathcal{U}_2} = \frac{1}{\mu_2} + \frac{1}{\lambda_2}. \tag{20}$$

V. Numerical results

Fig. 5 shows the average peak age ($\Delta_{peak,1}$) and the lower bound on the average age (Δ_{LB}) as computed in the previous section for stream $\mathcal{U}_1$ as well as the average age ($\Delta_{\mathcal{U}_2}$) of stream $\mathcal{U}_2$. In this plot, we fix $\mu_1 = 10$, $\mu_2 = 5$, $\lambda_1 = 2$ and we vary λ_2. As we can see, for stream $\mathcal{U}_1$ both the lower bound and the average peak age blow up when λ_2 gets close to $\frac{\mu_1\mu_2}{\lambda_1} - \mu_2$. This observation is in line with our result in Theorem 1 and the stability condition (3). It is easy to see via a coupling argument that if we increase λ_2, the age process $\Delta_{\mathcal{U}_1}(t)$ of the $\mathcal{U}_1$ stream will stochastically increase. We see from the plots that the lower bound on $\Delta_{\mathcal{U}_1}$ and its average peak exhibit the same behavior. On the other hand, the average age of stream $\mathcal{U}_2$ is decreasing in λ_2 (from (20)). Consequently, minimizing $\Delta_{\mathcal{U}_2}$ and minimizing $\Delta_{\mathcal{U}_1}$ are conflicting goals.

We have seen that the average age of stream $\mathcal{U}_2$ is not affected by the presence of the other stream. However, Fig. 5 shows the effect of stream $\mathcal{U}_2$ on the average age of stream $\mathcal{U}_1$. For that, we plot the average age (Δ_{ref}) of an M/M/1 queue with generation rate λ_1 and service rate μ_1 (given in [1]). We observe that the lower bound on the average age of stream $\mathcal{U}_1$ is very close to the value of the average age of this stream if stream $\mathcal{U}_2$ is not present. Only at high values of λ_2 (around 15) does Δ_{LB} clearly start to diverge from Δ_{ref}. However, $\Delta_{peak,1}$ starts diverging much sooner as expected from an upper bound. Nonetheless, for a non negligible interval of λ_2 (especially at low values), the upper bound $\Delta_{peak,1}$ and the lower bound Δ_{LB} are close and they give a tight approximation of the average age of stream $\mathcal{U}_1$. This means that at these values of λ_2, the average age achieved by stream $\mathcal{U}_1$ is very close to that achieved by the same stream with the absence of stream $\mathcal{U}_2$. Thus stream $\mathcal{U}_2$ has almost no effect on stream $\mathcal{U}_1$ for these values of λ_2 from an age point of view. The effect of stream $\mathcal{U}_2$ on the first one is especially felt at high values of λ_2, which is an expected behavior.

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. 66–70.

[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 queueing 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=EkSsgr-d8ZoC