二つの閾値がある FIFO 待ち行列の最悪値性能評価*

高田 寛之†

The Worst Case Evaluation for a FIFO Queue with Two Thresholds

by
Hiroyuki TAKADA*

We consider a loss system with a buffer and a FIFO server. There are a green flow and a yellow flow in the system. The system has two discarding thresholds corresponding to the types of flow. If an arriving packet looks at the buffer contents exceeding over the discarding threshold, then the packet is discarded. We take the fluid approach. Our aim is to present the upper bound of the partial queue length of the green flow.

Keywords: Network Calculus, Loss System, FIFO

1 Introduction

Various communication services have been able to run on TCP/IP networks by the development of computer and network technology and the spread of the Internet in recent years. In particular, real-time communication services require guaranteeing QoS.

DiffServ [1] standardized a framework guaranteeing QoS. Sato, Kobayashi, Pan, Tartarelli and Banchs (2001)[5] have provided configuration rule of DiffServe parameters. However, the study has not considered the increasing burstiness due to multiplexing[4].

Our issue is the worst case evaluation for the partial output burstiness (or almost equivalently the partial queue length) of the green flow in the discrete-time model introduced by [5]. We take a fluid approach for simplicity, so packetization is for feature works. Our computation is based on deterministic network calculus[3, 2].

The remainder of the paper is organized as follows: In Section 2, we introduce a loss FIFO system. The partial queue length of the green flow is defined in the section. Section 3 presents the upper bound of the partial queue length. Conclusion and feature works are given in Section 4.

2 Model

We consider a discrete-time queueing system. Any indexes of time are integers. For simplicity, $[s, t]$ represents a set of integers grater than or equal to s and lesser than or equal to t.

A token bucket filter receives packets from the external of the network. The filter colors them into two colors, green and yellow, which are labeled by G and

* 平成 22 年 12 月 15 日受理
† 情報システム工学科

Y, respectively. The green flow consists of the packets keeping the service level agreement (SLA). The yellow flow consists of the packets violating the SLA. Assume that the packets are instantaneously transferred from the filter to the dropper. For i = Y and G, $\overline{a}_i(t)$ denotes the amount of packets of type i arrives to the clipper on flow i at time t. $\overline{A}_i(s,t)$ is the cumulative function of sequence $\{\overline{a}_i(t)\}_{t\geq 1}$ in interval $(s,t]$, namely, $\overline{A}_i(s,t) = \sum_{u=s+1}^{t} \overline{a}_i(u)$.

The queueing system consists of a dropper on flow Y, a dropper on flow G, a buffer and a FIFO server. The dropper on flow i discards type i packets if the system is too crowded, it sends them to the queue otherwise. The crowded state is defined that $q(t) \geq \delta_i$, where $q(t)$ denotes the queue length at time t, which is the arrival time of the packets, and δ_i denotes the threshold of type i flow for i = Y and G. We suppose that $\delta_{\mathrm{Y}} \leq \delta_{\mathrm{G}}$. The discarding discipline is drop-tail. For simplicity, we take a fluid approach, The packets in the buffer are served in FIFO manner. If green packets and yellow packets arrive at the same time, then the yellow packets try to join earlier than any green packets. The assumption is required to determine the service order, because FIFO does not define the service order. Our service order gives the worst case for the green flow. The FIFO server is a work conserving server with a constant rate c. For $i = \mathrm{Y}, \mathrm{G}$, $a_i(t)$ denotes the amount of the survivals of type i at time t, namely, they are defined by

$$a_{\mathrm{Y}}(t) = \min(\overline{a}_{\mathrm{Y}}(t), (\delta_{\mathrm{Y}} + c - q(t-1))^+), \quad (1)$$
$$a_{\mathrm{G}}(t) = \min(\overline{a}_{\mathrm{G}}(t), (\delta_{\mathrm{G}} + c - q(t-1) - a_{\mathrm{Y}}(t))^+), \quad (2)$$

where $x^+ = \max(0, x)$. The amounts $(\delta_{\mathrm{Y}} + c - q(t-1))^+$ and $(\delta_{\mathrm{G}} + c - q(t-1) - a_{\mathrm{Y}}(t))^+$ represent the free space in the queue at time t. The loss of flow i at time t is given by $\overline{a}_i(t) - a_i(t)$. Let $a(t) = a_{\mathrm{Y}}(t) + a_{\mathrm{G}}(t)$ and $A(s,t) = \sum_{u=s+1}^{t} a(u)$. The buffer contents $q(t)$ is defined as $q(0) = 0$ and $q(t) = (a(t) + q(t-1) - c)^+$ for $t \geq 1$, or equivalently,

$$q(t) = \max_{u\in[0,t]}(A(u,t) - c \cdot (t-u)), \quad (3)$$

for $t \geq 0$.

In the view of Figure 1, the queueing system is a loss system, however, it is a lossless system in the view of Figure 2.

Let e_t denote the latest idle time before or just time t. Let w_t be the maximum value of arrival times of packets completely served until time t. w_t indicates the arrival time index of the packet in service at time t. Namely, they are defined by

$$e_t = \max\{u \in [0,t] \,|\, q(u) = 0\}, \quad (4)$$
$$w_t = \max\{v \in [e_t,t] \,|\, A(e_t,v) \leq c \cdot (t-e_t)\}. \quad (5)$$

It is easy to check $e_0 = w_0 = 0$ and $e_t \leq w_t \leq t$. In addition, $q(t) = 0$ if and only if $e_t = w_t = t$. Because $(e_t, t]$ is a busy period when $q(t) > 0$, the cumulative output in interval $(e_t, t]$ is given by $c \cdot (t - e_t)$. Since the system is a lossless and work conserving for the input process from the dropper, The cumulative input in $(e_t, t]$ must be equal to the summation of the cumulative output in $(e_t, t]$ and the queue length at time t, namely, $A(e_t, t) = c \cdot (t - e_t) + q(t)$ holds, or equivalently,

$$q(t) = A(e_t, t) - c \cdot (t - e_t). \quad (6)$$

$q(t) = 0$ is not contradicting (6) by $e_t = t$. On the other hand, w_t tells us the following inequalities.

$$A(w_t + 1, t) < q(t) \leq A(w_t, t), \quad (7)$$

Let $q_i(t)$ be the partial queue length of type i packets. We need to represent $q_{\mathrm{Y}}(t)$ and $q_{\mathrm{G}}(t)$ by the bivariate function A and c. We observe two scenarios in Example 1 to understand the partial queue length.

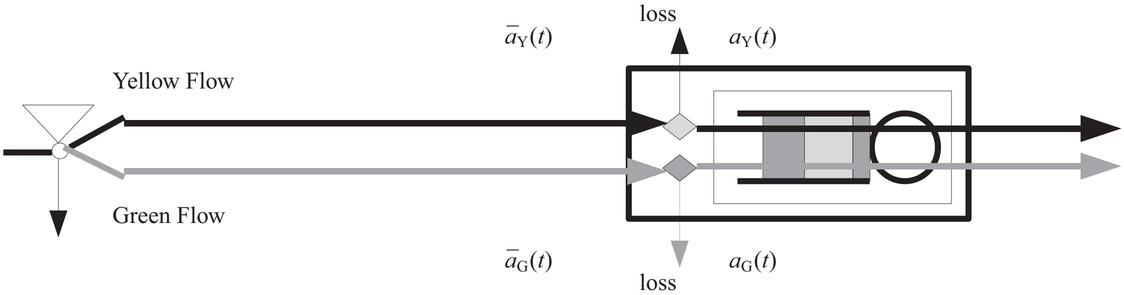

Figure.1 The loss system with two types of flow from a token bucket filter.

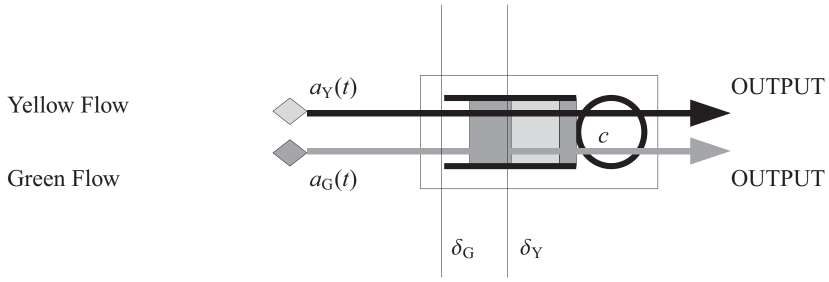

Figure.2 A part of the system is a lossless system.

Table.1 The input process of Scenario 1

t	1	2	3	4	5	6	7	8
$a_Y(t)$	1	1	0	2	4	3	2	4
$a_G(t)$	2	2	2	1	3	1	3	1
$a(t)$	3	3	2	3	7	4	5	5

Table.2 The input process of Scenario 2

t	1	2	3	4	5	6	7	8
$a_Y(t)$	1	1	0	2	3	4	2	4
$a_G(t)$	2	2	2	1	1	3	3	1
$a(t)$	3	3	2	3	4	7	5	5

Example 1. *Consider two scenarios given by Table 1 and Table 2 with $c = 2$, $\delta_Y = \delta_G = \infty$ and $t = 8$. Notice that $(a_Y(5), a_G(5), a(5))$ and $(a_Y(6), a_G(6), a(6))$ in Table 1 are swapped in Table 2. Because type Y packets enter earlier than tyep G packets if those arrival times are same, the service order is alternative (see Figure 1 and Figure 2).*

Time 0 is the last idle time, i.e., $e_t = e_8 = 0$. In addition, $w_t = w_8 = 4$ in scenario 1 and $w_t = w_8 = 5$ in scenario 2. Since $A(0, 8) = 32$ and $c \cdot (t - e_t) = 16$, $q(t) = q(8) = 32 - 16 = 8$. Since $q_Y(t)$ and $q_G(t)$ are the amount of type i packet in the queue, We obtain that $q_Y(8) = 9$ and $q_G(8) = 7$ in both cases by counting the amounts of packets. It is coincidence that the partial queue lengths of Scenario 1 equal to ones of Scenario 2.

For generality of counting the partial queue lengths, notice that the queue consists of the packets arrived $(w_t + 1, t]$ and ones arrived at $w_t + 1$. Any packets arrived $(w_t + 1, t]$ are still in the queue at time t. If

$$A(e_t, w_t) + a_Y(w_t + 1) \leq c \cdot (t - e_t),$$

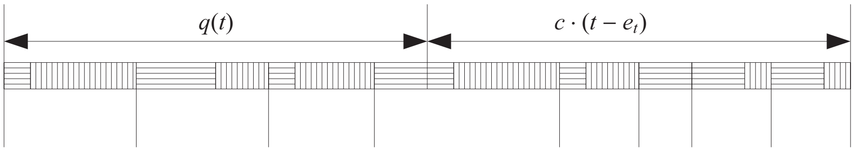

Figure.3　The service order of Scenario 1. The vertical and horizontal line boxes are type Y and type G packets, respectively.

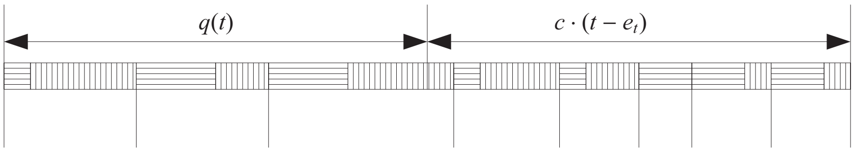

Figure.4　The service order of Scenario 2. The vertical and horizontal line boxes are type Y and type G packets, respectively.

(see also Figure 3), then the size of type Y packets arrived at $w_t + 1$ in the queue is 0, besides the size of type G packets arrived at $w_t + 1$ in the queue is

$$a_{\mathrm{G}}(w_t + 1) - [c \cdot (t - e_t) - A(e_t, w_t) - a_{\mathrm{Y}}(w_t + 1)].$$

Otherwise, namely,

$$A(e_t, w_t) + a_{\mathrm{Y}}(w_t + 1) \geq c \cdot (t - e_t),$$

(see also Figure 4), the sizes are

$$A(e_t, w_t) + a_{\mathrm{Y}}(w_t + 1) - c \cdot (t - e_t)$$

and $a_{\mathrm{G}}(w_t + 1)$, respectively. Hence, $q_{\mathrm{Y}}(t)$ is given by

$$\begin{aligned} & A_{\mathrm{Y}}(w_t + 1, t) \\ & \qquad + (A(e_t, w_t) + a_{\mathrm{Y}}(w_t + 1) - c \cdot (t - e_t))^+ \qquad (8) \\ & = \max\,[A_{\mathrm{Y}}(w_t + 1, t), \\ & \qquad A_{\mathrm{Y}}(e_t, t) + A_{\mathrm{G}}(e_t, w_t) - c \cdot (t - e_t)]. \end{aligned}$$

Furthermore, $q_{\mathrm{G}}(t)$ is given by

$$\begin{aligned} & A_{\mathrm{G}}(w_t + 1, t) \\ & \qquad + \min(a_{\mathrm{G}}(w_t + 1), A(e_t, w_t + 1) - c \cdot (t - e_t)) \\ & = A_{\mathrm{G}}(w_t, t) \\ & \qquad - (c \cdot (t - e_t) - A(e_t, w_t) - a_{\mathrm{Y}}(w_t + 1))^+ \qquad (9) \\ & = \min\,[A_{\mathrm{G}}(w_t, t) \\ & \qquad A_{\mathrm{G}}(e_t, t) + A_{\mathrm{Y}}(e_t, w_t + 1) - c \cdot (t - e_t)]. \qquad (10) \end{aligned}$$

3　Maximum Partial Queue length and Envelope of Yellow flow

Theorem 1. *Assume that*

$$A_G(s, t) \leq r \cdot (t - s) + b - r, \quad \leq \forall s \leq \forall t.$$

Then the partial queue length of the green flow is bounded as follows:

$$q_G(t) \leq b + (r/c) * \delta_Y, \quad \forall t \geq 0.$$

Before the proof of Theorem 1, we require the envelope function on the yellow flow.

Lemma 1. *We have*

$$A_Y(u, t) \leq \max_{\tau \in [u, t-1]} (-A_G(u, \tau) + c \cdot (\tau - u + 1) + \delta_Y). \tag{11}$$

Proof. (1) yields

$$\begin{aligned} & a_{\mathrm{Y}}(t) \\ & \leq (c + \delta_{\mathrm{Y}} - q(t-1))^+ \\ & = \left[c + \delta_{\mathrm{Y}} - \max_{u \in [0, t-1]} (A(u, t-1) - c \cdot (t - 1 - u))\right]^+ \\ & = \min_{u \in [0, t-1]} (-A(u, t-1) + c \cdot (t - u) + \delta_{\mathrm{Y}})^+, \end{aligned}$$

or equivalently,

$$a_{\mathrm{Y}}(t) \leq (-A(u, t-1) + c \cdot (t - u) + \delta_{\mathrm{Y}})^+,$$

for any $u \in [0, t-1]$. Adding $A_{\mathrm{Y}}(u, t-1)$ to the both sides, we obtain the recursive inequality

$$A_{\mathrm{Y}}(u,t) \le \max\left[A_{\mathrm{Y}}(u,t-1),\right. \\ \left. -A_{\mathrm{G}}(u,t-1) + c\cdot(t-u) + \delta_{\mathrm{Y}}\right], \quad (12)$$

for $u \in [0, t-1]$. Substituting (12) to itself recursively, we obtain (11). Namely,

$$\begin{aligned} A_{\mathrm{Y}}(u,t) &\le \max\left[A_{\mathrm{Y}}(u,t-2),\right. \\ &\quad -A_{\mathrm{G}}(u,t-2) + c\cdot(t-1-u) + \delta_{\mathrm{Y}}, \\ &\quad \left. -A_{\mathrm{G}}(u,t-1) + c\cdot(t-u) + \delta_{\mathrm{Y}}\right] \\ &\le \cdots \\ &\le \max\left[A_{\mathrm{Y}}(u,u),\right. \\ &\quad -A_{\mathrm{G}}(u,u) + c\cdot(u+1-u) + \delta_{\mathrm{Y}}, \\ &\quad -A_{\mathrm{G}}(u,u+1) + c\cdot(u+2-u) + \delta_{\mathrm{Y}}, \\ &\quad \ldots \\ &\quad -A_{\mathrm{G}}(u,t-2) + c\cdot(t-1-u) + \delta_{\mathrm{Y}}, \\ &\quad \left. -A_{\mathrm{G}}(u,t-1) + c\cdot(t-u) + \delta_{\mathrm{Y}}\right] \\ &= \max_{\tau\in[u,t-1]} (-A_{\mathrm{G}}(u,\tau) + c\cdot(\tau+1-u) + \delta_{\mathrm{Y}}), \end{aligned}$$

where, in the last equation, $\max(A_{\mathrm{Y}}(u,u), c+\delta_{\mathrm{Y}}) = c + \delta_{\mathrm{Y}}$ is used. □

We are ready to show Theorem 1.

Proof. of Theorem 1. We have from (4) and (5) that $0 \le e_t \le w_t \le t$. From the inequalities, (10) yields

$$q_{\mathrm{G}}(t) \le \max_{v\in[0,t]} \max_{u\in[0,v]} \min\left[A_{\mathrm{G}}(v,t),\right. \\ \left. A_{\mathrm{G}}(u,t) + A_{\mathrm{Y}}(u,v+1) - c\cdot(t-u)\right]. \quad (13)$$

Applying Lemma 1 to the right hand side of (13), $q_{\mathrm{G}}(t)$ is evaluated by

$$\begin{aligned} &q_{\mathrm{G}}(t) \\ &\le \max_{v\in[0,t]} \max_{u\in[0,v]} \min\left[A_{\mathrm{G}}(v,t), A_{\mathrm{G}}(u,t) - c\cdot(t-u)\right. \\ &\quad \left. + \max_{\tau\in[u,v]}(-A_{\mathrm{G}}(u,\tau) + c\cdot(\tau-u+1) + \delta_{\mathrm{Y}})\right] \\ &= \max_{v\in[0,t]} \max_{u\in[0,v]} \min\left[A_{\mathrm{G}}(v,t),\right. \\ &\qquad \left. \max_{\tau\in[u,v]} A_{\mathrm{G}}(\tau,t) - c\cdot(t-\tau) + c + \delta_{\mathrm{Y}}\right]. \end{aligned}$$

Let $\Psi(\tau,t) = A_{\mathrm{G}}(\tau,t) - c\cdot(t-\tau) + c + \delta_{\mathrm{Y}}$. Since $\max_{\tau\in[u,v]} \Psi(\tau,t)$ is nonincreasing on u, the maximum value in $u \in [0,v]$ is $\max_{\tau\in[0,v]} \Psi(\tau,t)$. So,

$$\begin{aligned} q_{\mathrm{G}}(t) &\le \max_{v\in[0,t]} \min\left[A_{\mathrm{G}}(v,t), \max_{\tau\in[0,v]} \Psi(\tau,t)\right] \\ &= \max_{v\in[0,t]} \max_{\tau\in[0,v]} \min\left[A_{\mathrm{G}}(v,t), \Psi(\tau,t)\right] \\ &= \max_{\tau\in[0,t]} \max_{v\in[\tau,t]} \min\left[A_{\mathrm{G}}(v,t), \Psi(\tau,t)\right] \\ &= \max_{\tau\in[0,t]} \min\left[\max_{v\in[\tau,t]} A_{\mathrm{G}}(v,t), \Psi(\tau,t)\right] \\ &= \max_{\tau\in[0,t]} \min\left[A_{\mathrm{G}}(\tau,t), \Psi(\tau,t)\right]. \end{aligned}$$

Since $\min(A_{\mathrm{G}}(\tau,t), \Psi(\tau,t)) = A_{\mathrm{G}}(\tau,t) + \min(-c\cdot(t-\tau) + c + \delta_{\mathrm{Y}})$, we have

$$q_{\mathrm{G}}(t) \le \max_{\tau\in[0,t]} \left[A_{\mathrm{G}}(\tau,t) - [c\cdot(t-\tau) - c - \delta_{\mathrm{Y}}]^+\right] \quad (14)$$

In addition, we apply inequality $A_{\mathrm{G}}(\tau,t) \le r\cdot(t-\tau) + b - r$ to (14).

$$\begin{aligned} &q_{\mathrm{G}}(t) \\ &\le \max_{\tau\in[0,t]} \left[r\cdot(t-\tau) + b - r - [c\cdot(t-\tau) - c - \delta_{\mathrm{Y}}]^+\right] \\ &= \max_{x\in[0,t]} \min\left[rx + b - r, (r-c)x + b - r + c + \delta_{\mathrm{Y}}\right] \\ &= r\cdot\left(1 + \frac{\delta_{\mathrm{Y}}}{c}\right) + b - r \\ &= b + \frac{r}{c}\delta_{\mathrm{Y}}. \end{aligned}$$

Thus, the proof completes. □

The right hand side of (14) is represented as $(A_{\mathrm{G}} \oslash S_{\mathrm{G}})(t,t)$ in terms of network calculus where the deconvolution operator $\oslash$ between bivariate functions F and G is defined by

$$(F \oslash G)(s,t) = \max_{u\in[0,s]} [F(u,t) - G(u,s)],$$

besides $S_{\mathrm{G}}(s,t)$ is given by

$$S_{\mathrm{G}}(s,t) = [c\cdot(t-s) - c - \delta_{\mathrm{Y}}]^+.$$

We should consider the aggregate flow for the delay evaluation, because delay effects both flows. In the same way, the lossless condition is determined by the total queue length, because $q_{\mathrm{G}}(t) \le q(t) \le \delta_{\mathrm{G}}$. Strictly, we should evaluate the amount of loss $\overline{a}_{\mathrm{G}}(t) - a_{\mathrm{G}}(t)$.

4 Conclusion and feature works

We studied a loss FIFO system with two drop-level inputs introduced by [5]. By the theory of deterministic network calculus, we obtained the envelope function on the yellow flow and the upper bound of the partial queue length.

We did not give the envelope function of the raw input $\overline{A}$. The clipper on the yellow flow gave another envelope of the modified input $\overline{A}_{\mathrm{Y}}$.

We can assume the envelope function of the raw input. If the flows are colored by a double token buckets mechanism, then the situation is natural. However, the computation is hard very much.

Using the theory of stochastic network calculus, we guess that we can evaluate the upper bound of the loss probability with the significant level α percent.

References:

1) S. Blake, D. Black, M. Carlson, E. Davies (1998) An Architecture for Differentiated Services, *RFC 2475*. www.ietf.org/rfc/rfc2475.txt

2) J. Y. L. Boudec and P. Thiran (2004) *NETWORK CALCULUS: A Theory of Deterministic Queueing Systems for the Internet*, Springer.

3) C. S. Chang (2000) *Performance Guarantees in Communication Networks*, Springer.

4) V. Cholvia, J. Echagüe, J. Y. Le Boudec (2002) Worst case burstiness increase due to FIFO multiplexing, *Performance Evaluation*, Vol.49, 491-506.

5) S. Sato, K. Kobayashi, H. Pan, S. Tartarelli, A. Banchs (2001) Configuration rule and performance evaluation for DiffServ parameters, *Proc.* of ITC17, 931-942.