Many Sources Asymptotics for a Feedforward Network with Small Buffers *

Ozcan Ozturk
Dept. of Electrical Engineering
Purdue University
West Lafayette IN 47906
ozturk@purdue.edu

Ravi R. Mazumdar
Dept. of Electrical Engineering
Purdue University
West Lafayette IN 47906
mazum@ecn.purdue.edu

Nikolay LIKHANOV
Institute for Problems of Information Transmission
Russian Academy of Sciences, Moscow, Russia

Abstract

In this paper we obtain the logarithmic asymptotics in a network of small buffers when the resources are accessed by a large number of stationary independent sources. Under the assumption that the network is feedforward with respect to source-destination pairs we identify the precise large deviations rate functions at each node in terms of the external input characteristics. Using these results we then show that the asymptotic admissible region when each type of source requires the same QoS in terms of loss is the same as that which is obtained by assuming that flows pass through each node unchanged.

1 Introduction

One of the challenging problems in networks is to characterize the admissible region of the numbers of connections or flows that can be admitted into the network in order to guarantee a given level of Quality of Service (QoS). Deterministic approaches based on the so-called *network calculus* [2, 1] approach are powerful in that we can treat the end-to-end problem but yield conservative results due to the fact that it is essentially a worst-case approach. Providing statistical QoS is much more efficient but the analysis is much more complicated. This is due to the fact that streams undergo changes in their statistics when "filtered" through queues and precise characterization of the statistical behavior is possible only in a few simple cases.

In this paper we consider the situation in a network where a large number of flows are involved between origin-destination pairs. In particular, such an architecture is close to the MPLS case where virtual pipes are established for connections which we assumed are identified by their "routes". We assume that individual flows are independent when they arrive into the network. Our interest is to determine the admissible region (of allowable flows) in order to provide a statistical QoS based on keeping the loss probability small (below a given value $\varepsilon << 1$). Noting that losses add across the buffers the flows pass is

*This research was supported in part by NSF Grants No. 0099137-ANI and 0087404-ANI

a very important structural fact since these depend on the tail distributions of the buffer occupancy in the buffers encountered by each route. Given the fact that ε is small, we can assume that we are in the region of large deviations. This leads us naturally to characterize the asymptotics of the buffer occupancy distribution.

In this paper, we investigate the many sources asymptotics of a network with small buffers. Many sources asymptotics for a single node have been investigated quite thoroughly [5, 7, 3]. It has been shown that the buffer overflow probability goes to zero and this happens most likely in a busy period of a particular duration corresponding to a "time scale". There have been attempts to generalize these results to the network case. Wischik [8] has shown that the m.g.f. for a single traffic does not change as it passes through a node when multiplexed with many similar inputs. Recently, Eun and Shroff [4] have shown that in a two-stage queueing system where the first node serves many flows, of which a certain fixed finite set arrive to the second node; the first node can be ignored for calculating the overflow probability of the second node as the number of flows of the first node (N) increases. It is difficult to get similar results when the number of output flows arriving at the downstream node increases with N when the buffers are large (scaled with N). This situation is however of interest when trying to estimate the end-to-end QoS in the general scenario described in the beginning. Although overflow probability still goes to zero at the downstream node, determining the rate of this decay depends on the particular time scale and characteristics of the output flows from the upstream nodes, which in turn depend on the sample paths of the external inputs.

When the buffers are small, the above mentioned time scale is "1" in discrete time (and goes to zero in continuous time) at all nodes and thus the overflow probability asymptotics depend on only the instantaneous rates of the traffic. By exploiting this, we find the overflow probability asymptotics in terms of the large deviation (LD) rate functions of the inputs. We then define the packet loss rates and find their LD rates along with acceptance region. When all of the input traffics require the same QoS in terms of packet loss rate we show that the upstream nodes can be ignored while determining the acceptance region. For bounded traffics, a lower bound for the their LD rate functions (and hence an upper bound for the overflow probabilities) can be easily calculated.

This small buffer scenario is actually of much interest in today's networks where buffers are small. This is the essence of the so-called **rate envelope multiplexing** in networks (see [6]) where buffers are small to absorb local fluctuations but essentially the network can be modelled by bufferless nodes.

The outline of this paper is as follows: in Section 2 we obtain the LD asymptotics for the buffer occupancy. In Section 3 we determine the asymptotics of the loss process and then determine the admissible region. In Section 4 we discuss the results and provide some approximations in cases when only a coarse statistical characterization of the traffic can be given.

2 Model and Large Deviations of Workload

Consider a feedforward network of K nodes which is accessed by M types of independent traffics (flows). We consider a discrete time fluid FIFO model where traffic arrives at times $t \in Z$ and served immediately. Stationary and ergodic input traffic of type $m = 1, \ldots, M$, denoted by $X^{m,N}$, has rate $X_t^{m,N}$ at time t. Let $\rho_m = E[X_0^{m,N}]/N$ and $X^{m,N}(t_1, t_2) = \sum_{t=t_1}^{t_2-1} X_t^{m,N}$. We will assume that $X^{m,N}(0,t)/N$ has a large deviation principle (LDP) with the rate function $I_t^{X^m}(x)$. In the remainder, we will drop the superscript N of $X^{m,N}$

and also write $I^{X^m} \equiv I_0^{X^m}$. Also, we assume the following condition which is satisfied for all the known traffic models including long range traffic.

$$\text{For any m and } a > \rho_m, \quad \liminf_{t\to\infty} \frac{I_t^{X^m}(at)}{\log t} > 0.$$

Node $k = 1, \ldots, K$ has output rate NC_k and buffer capacity $B_k(N)$ where $B_k(N)/N \to 0$. Type m traffic has a fixed route and its path is represented by the vector $\mathbf{k}^m = \{k_1^m, \ldots, k_{l_m}^m\}$ where $k_i^m \in \{1, \ldots, K\}$. Hence, type m traffic traverses the nodes $\{k_i^m\}$ by entering the network at node k_1^m and leaving after node $k_{l_m}^m$. For each node k, define the set of traffic types which pass through node k by $\mathcal{M}^k = \{m : \ k_i^m = k, \ 1 \le i \le l_m\}$.

Define $X_t^{m,k}$ ($Y_t^{m,k}$) be the input (output) rate at time t corresponding to the traffic of type m at the node k . If node k is not on the path of input m, then $X_t^{m,k} = Y_t^{m,k} = 0$.

2.1 Two Node Case

We first analyze the case when the network is composed of just two nodes in tandem. The buffer asymptotics at the first node follows from the single node results in [3, 5].

For $n = 1, \ldots, M$, define the function $f_n : \mathbb{R}^{M+1} \to \mathbb{R}$ as

$$f_n(x_1, \ldots, x_{M+1}) = \frac{x_n x_{M+1}}{\max(\sum_{i=1}^M x_i, x_{M+1})}$$

Also, for $x = (x_m) \in \mathbb{R}^M$, define $\bar{x} \in \mathbb{R}^M$ such that

$$\bar{x}_m = \begin{cases} x_m & m \in \mathcal{M}^1 \\ 0 & otherwise \end{cases}$$

Proposition 1. *Let Q_t^2 be the stationary buffer content of node 2 at time t. Then,*

$$\lim_{N\to\infty} \tfrac{1}{N} \log \mathbb{P}(Q_0^2 > 0) =$$

$$- \inf\{\textstyle\sum_{m=1}^M I^{X^m}(x_m) \mid x = (x_m) \in \mathbb{R}^M, \sum_{m\in\mathcal{M}^1 \bigcap \mathcal{M}^2} f_m(\bar{x}, C_1) + \sum_{m\in\mathcal{M}^2-\mathcal{M}^1} x_m = C_2\} \tag{1}$$

Proof. Since the service is FIFO, at time $t = 0$, the output capacity available for type $m \in \mathcal{M}^1$ traffic at node 1 is proportional to the ratio of X_0^m to the total input to node 1 at time 0 after the traffic in the buffer is served. Then,

$$f_m(X_0^{1,1}, \ldots, X_0^{M,1}, NC_1 - B_1(N)) \le Y_0^{m,1} \le f_m(X_0^{1,1}, \ldots, X_0^{M,1}, NC_1) + B_1(N) \tag{2}$$

By definition, $X_0^{m,1} = X_0^m$ if $m \in \mathcal{M}^1$ and 0 otherwise. Note that $f_m(X_0^{1,1}, \ldots, X_0^{M,1}, Na)/X_0^{m,1}$ is the amount of capacity of node 1 used by $X_0^{m,1}$ if total capacity available was Na. Since $f_m(X_0^{1,1}, \ldots, X_0^{M,1}, NC_1 - B_1(N)) \ge f_m(X_0^{1,1}, \ldots, X_0^{M,1}, NC_1) - B_1(N)$, from the contraction principle, $Y_0^{m,1}/N$ satisfies LDP with the rate function given by

$$I^{Y^{m,1}}(y) = \inf\{\sum_{m=1}^M I^{X^m}(x_m) \mid x = (x_m) \in \mathbb{R}^M, f_m(\bar{x}, C_1) = y\}$$

Let Z_t^2 be the rate of total input traffic at node 2 at time t, i.e.,

$$Z_t^2 = \sum_{m=1}^{M} X_t^{m,2}$$

and $Z^2(t_1, t_2) = \sum_{t=t_1}^{t_2-1} Z_t^2$. Then, Z_0^2/N satisfies LDP with rate function

$$I^{Z^2}(y) = \inf\left\{\sum_{m=1}^{M} I^{X^m}(x_m) \mid x = (x_m) \in \mathbb{R}^M, \sum_{m\in\mathcal{M}^1\bigcap\mathcal{M}^2} f_m(\bar{x}, C_1) + \sum_{m\in\mathcal{M}^2-\mathcal{M}^1} x_m = y\right\}$$

Now we find the LDP rate function for the buffer content. First, note that

$$Q_0^2 \leq \sup_{t\geq 0} Z^2(-t,0) - NC_2 t$$

Also,

$$Z^2(t_1,t_2) \leq \sum_{m\in\mathcal{M}^2} X^m(t_1,t_2) + (t_2-t_1)B_1(N)$$

Then,

$$\mathbb{P}(Q_0^2 > 0) \leq \sum_{t=1}^{\infty} \mathbb{P}(Z^2(-t,0) > NC_2 t) \leq$$

$$\sum_{t=1}^{t_0} t\mathbb{P}(Z_0^2 > NC_2) + \sum_{t=t_0+1}^{\infty} \mathbb{P}(Z^2(-t,0) > NC_2 t) \leq$$

$$\sum_{t=1}^{t_0} t\mathbb{P}(Z_0^2 > NC_2) + \sum_{k=t_0+1}^{\infty} \mathbb{P}(\sum_{m\in\mathcal{M}^k} X^m(-t,0) + tB_1(N) > NC_2 t)\}$$

Choose a_m such that $a_m > \rho_m$ and $\sum_{m\in\mathcal{M}^2} a_m < C_2$. Furthermore, take t_0 large enough so that for all $t > t_0$ and $m \in \mathcal{M}^2$, $I_t^{X^m}(a_m t) > \alpha \log t > I^{Z^2}(C_2)$ for some $\alpha > 0$ and also $\sum_{m\in\mathcal{M}^2} a_m + \frac{B_1(N)}{N} < C_2$ for large enough N. Then,

$$\mathbb{P}(Q_0^2 > 0) \leq \sum_{t=1}^{t_0} t\mathbb{P}(Z_0^2 > NC_2) + |\mathcal{M}^2| \frac{e^{-N\alpha\log t_0}}{N\alpha - 1}$$

and hence

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Q_0^2 > 0) \leq \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Z_0^2 > C_2) = -I^{Z^2}(C_2)$$

Now, we look at the lower bound. First,

$$\mathbb{P}(Q_0^2 > 0) \geq \mathbb{P}(Z_0^2 > NC_2)$$

and thus

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Q_0^2 > 0) \geq \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Z_0^2 > NC_2) = -I^{Z^2}(C_2)$$

which completes the proof.

□

2.2 General Case

First result is an extension of the above.

Proposition 2. *There exists a function $g^{m,k} : \mathbb{R}^M \to R$, relating the instantaneous input rate at node k for traffic m to all of the instantaneous input traffic rates such that*

$$X_0^{m,k}/N = g^{m,k}(X_0^1/N, \ldots, X_0^M/N) + o(N) \tag{3}$$

Let Q_t^k be the stationary buffer content of node k at t. Then,

$$\begin{aligned} &\lim_{N\to\infty} \tfrac{1}{N} \log \mathbb{P}(Q^k > 0) = -I_k \doteq \\ &-\inf\{\textstyle\sum_{m=1}^M I^{X^m}(x_m) \mid x = (x_m) \in \mathbb{R}^M, \sum_{m=1}^M g^{m,k}(x) = C_k\} \end{aligned} \tag{4}$$

The proof depends on using induction on the nodes to show the existence of (3). To this end, a degree function on the nodes is defined as follows:

$$deg(node\ k) = \max\{i : k_i^m = k,\ i = 1, \ldots, l_m,\ m \in \mathcal{M}^k\}$$

Hence, the degree of a node is the maximum number of upstream nodes passed by any source before accessing this particular node. This is well defined for any node k since it does not appear more than once in the path of any traffic. The induction step at $deg(node\ k) = j + 1$ mimics the steps in the two node case. But now, the function $g^{m,k}$ is defined recursively and hard to be expressed explicitly.

We can also find the joint distribution of workloads by using the vector version of the contraction principle:

Proposition 3. *For any set $\mathcal{K} \subseteq \{1, \ldots, K\}$,*

$$\begin{aligned} &\lim_{N\to\infty} \tfrac{1}{N} \log \mathbb{P}(Q_0^k > 0, k \in \mathcal{K}) = \\ &-\inf\{\textstyle\sum_{m=1}^M I^{X^m}(x_m) \mid x = (x_m) \in \mathbb{R}^M, \sum_{m=1}^M g^{m,k}(x) = C_k, k \in \mathcal{K}\} \end{aligned}$$

3 Loss Ratio and Acceptance Region

We now consider the asymptotics of the loss traffic due to overflow. Assume that the following tail condition is satisfied for the input traffic:

$$\lim_{y\to\infty} \limsup_{N\to\infty} \frac{1}{N} \log E[X_0^m 1_{\{X_0^m > Ny\}}] = -\infty. \tag{5}$$

It can easily be shown that this condition is satisfied when X^m is the sum of N iid processes. Let L^m be the total packet loss ratio (PLR) , defined as the ratio of the expected value of lost traffic to the mean of input traffic (see below), for X^m. Then,

Proposition 4.

$$\lim_{N\to\infty} \frac{1}{N} \log L^m = -\min_{i=1,\ldots,l_m} I_{k_i^m}$$

Proof. Let $L^{m,k}$ be the expected value of cell loss at node k for type m traffic, i.e.

$$L^{m,k} = E[\max(0, X_t^{m,k} + Q_{t-1}^{k,m} - C_t^{k,m} - Q_t^{k,m})]$$

where $Q_t^{k,m}$ and $C_t^{k,m}$ are the buffer space and output capacity allocated for type m traffic at node k and time t. Then,

$$\log L^m = \log(\sum_{i=1}^{l_m} L^{m,k_i^m}) - \log(E[X_t^m]).$$

Note that, at a node $k = k_i^m$,

$$\max(0, X_0^{m,k} + Q_{-1}^{k,m} - C_0^{k,m} - Q_0^{k,m}) \leq X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k - B_k(N)) \quad (6)$$

and

$$\max(0, X_0^{m,k} + Q_{-1}^{k,m} - C_0^{k,m} - Q_0^{k,m}) \geq X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) - B_k(N) \quad (7)$$

Hence,

$$E[\max(0, X_t^{m,k} + Q_{t-1}^{k,m} - C_t^{k,m} - Q_t^{k,m})] \leq E[X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k - B_k(N))]$$

$$\leq Ny\mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k - B_k(N)) > 0) + E[X_0^{m,k} 1_{\{X_0^{m,k} > Ny\}}]$$

$$\leq Ny\mathbb{P}(Z_0^k > NC - B_k(N)) + E[X_0^{m,k} 1_{\{X_0^{m,k} > Ny\}}]$$

Now choose y such that

$$\limsup_{N\to\infty} \frac{1}{N} \log E[X_0^{m,k} 1_{\{X_0^{m,k} > Ny\}}] < -I_k$$

This is possible because of the tail condition (5) and $X_0^{m,k} \leq X_0^m + o(N)$. From

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Z_0^k > NC - B_k(N)) = -I_k$$

we get

$$\limsup_{N\to\infty} \frac{1}{N} \log L^{m,k} \leq -I_k$$

Adding up $L^{m,k}$ and $\lim_{N\to\infty} \frac{1}{N} \log(E[X_t^m]) = 0$ gives

$$\limsup_{N\to\infty} \frac{1}{N} \log L^m \leq \min_{i=1,\ldots,l_m} I_{k_i^m}$$

Now we find the lower bound for L^m. Let k be the node with the smallest i index where $\min_{i=1,\ldots,l_m} I_{k_i^m}$ is achieved and $k = k_{i_0}^m$ Then, from (7),

$$E[\max(0, X_t^{m,k} + Q_{t-1}^{k,m} - C_t^{k,m} - Q_t^{k,m})] \geq E[X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) - B_k(N)]$$

$$\geq Ny\mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) > B_k(N) + Ny)$$

$$\geq Ny\mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) > 2Ny)$$

where $y > 0$ and N is large enough to make $Ny > B_k(N)$.

Choose a, b such that $\frac{ab}{C_k+b} > 2y$. Then,

$$\mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) > 2Ny) \geq \mathbb{P}(X_0^{m,k} > Na, Z_0^k > NC_k + Nb)$$

$$= \mathbb{P}(Z_0^k > NC_k + Nb) - \mathbb{P}(X_0^{m,k} < Na, Z_0^k > NC_k + Nb)$$

Next,

$$\mathbb{P}(X_0^{m,k} < Na, Z_0^k > NC_k + Nb) \leq$$

$$\mathbb{P}(X_0^{m,k} < Na, Z_0^k > NC_k + Nb, Q^j = 0, j = k_1^m, \ldots, k_{(i_0-1)}^m) + \sum_{j=k_1^m,\ldots,k_{(i_0-1)}^m} \mathbb{P}(Q^j > 0)$$

By the choice of the node k, it is obvious that

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Q^j > 0) > \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(Z_0^{m,k} > NC_k)$$

for $j = k_1^m, \ldots, k_{(i_0-1)}^m$. Also note that if $Q^j = 0, j = k_1^m, \ldots, k_{(i_0-1)}^m$ holds, then all the traffic pass through without loss or delay at these nodes and hence $X_0^{n,k}$, $n \neq m$ does not depend on X_0^m. Therefore, if we choose,for example, $a = \rho_m/2$, then

$$\mathbb{P}(X_0^{m,k} < Na, Z_0^k > NC_k + Nb, Q^j = 0, j = k_1^m, \ldots, k_{(i_0-1)}^m)$$

$$\preceq \mathbb{P}(X_0^{m,k} < N\rho_m, Z_0^k > NC_k + Nb, Q^j = 0, j = k_1^m, \ldots, k_{(i_0-1)}^m) \leq \mathbb{P}(Z_0^k > NC_k + Nb)$$

Thus, combining above and taking $a = \rho_m/2,\ y < \rho_m/8$

$$\mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) > 2Ny) \geq \mathbb{P}(Z_0^k > NC_k + 8NC_k y/\rho_m)$$

and hence

$$\liminf_{N\to\infty} \tfrac{1}{N} \log \mathbb{P}(X_0^{m,k} - f_m(X_0^{1,k}, \ldots, X_0^{M,k}, NC_k) > B_k(N) + Ny)$$

$$\geq \liminf_{N\to\infty} \tfrac{1}{N} \log \mathbb{P}(Z_0^k > NC_k + 8NC_k y/\rho_m) = -I^{Z^k}(C_k + 8C_k y/\rho_m)$$

Now, letting $y \to 0$ gives

$$\liminf_{N\to\infty} \frac{1}{N} \log L^{m,k} \geq -I_k$$

and hence

$$\lim_{N\to\infty} \frac{1}{N} \log L^m = -\min_{i=1,\ldots,l_m} I_{k_i^m}$$

□

Now we assume that X^m is the sum of Nn_m i.i.d processes.

Corollary 1. *Let D be the acceptance region for (n_m) satisfying $\lim_{N\to\infty} \frac{1}{N} \log L^m < -\gamma_m$ for some $\gamma_m > 0$. Consider the fictional system where X^m goes to a node on its path without being affected by the previous nodes and let $\bar{D}$ be the acceptance region for this case. Then*

$$\bar{D} \subseteq D$$

Furthermore, if $\gamma_m = \gamma$, then

$$\bar{D} = D$$

Proof. For the fictional system, let $\bar{Q}^k$ be the stationary buffer content at node k and $t = 0$ and $\bar{I}_k = \lim_{N\to\infty} \frac{1}{N} \log P(\bar{Q}^k > B_k(N)$. Since

$$Z_0^k \leq \sum_{m\in\mathcal{M}^k} X_0^m + \sum_{k=1}^{K} B_k(N)$$

it follows that $\bar{Q}^k \geq Q^k$ and hence $\bar{I}_k \leq I_k$. This implies that $\bar{D} \subseteq D$.

If $\gamma_m = \gamma$ for all m, then note that

$$D = \{(n_m) : I_k > \gamma\} \text{ and } \bar{D} = \{(n_m) : \bar{I}_k > \gamma\}$$

Take $(n_m) \in D$. Then, $I_k > \gamma$ for every $k = 1, \ldots, K$. By definition, $\bar{I}_k = \sum_{m\in\mathcal{M}^k} I^{X^m}(\bar{x}_m)$ for some $\bar{x}_m$. Take $x \in \mathbb{R}^M$ as

$$x_m = \begin{cases} \bar{x}_m & m \in \mathcal{M}^k \\ \rho_m & \textit{otherwise} \end{cases}$$

Assume $g^{m,k}(x) < x_m$ for some $m \in \mathcal{M}^k$. Then, from the way $g^{m,k}$ has been defined, there must exist a node k' for which $\sum_{m\in\mathcal{M}^{k'}} g^{m,k'}(x) > C_{k'}$. Otherwise, the instantaneous rates would pass through the nodes without any change. Therefore, $\sum_{m=1}^{M} I^{X^m}(x_m) = \bar{I}_k > I_{k'}$. Since $(n_m) \in D$, $I_{k'} > \gamma$ and hence $\bar{I}_k > \gamma$.

If $g^{m,k}(x) = x_m$ for all $m \in \mathcal{M}^k$, then $\sum_{m\in\mathcal{M}^k} g^{m,k}(x) > C_k$ and thus $\bar{I}_k \geq I_k > \gamma$. Therefore, (n_m) is also in D', which completes the proof.

□

Remark 1. *Above corollary shows that when QoS (in terms of cell loss) is same for all the sources, we can easily find the acceptance region by considering the statistics of the sources as they enter the network and assuming that they do not change while passing through a node.*

4 Discussion

When there is traffic entering the network with total rate $o(N)$, above results do not change. The large deviation rate function for these small sources will be 0 at 0 and infinite elsewhere and hence they do not effect the minimization process in finding the asymptotics of the buffers.

If the buffers are large so that $B_k(N) \geq O(N)$, there is a time scale to overflow and it is difficult to find the large deviation rate function of an output in terms of input at time scales bigger than 1.

Continuous time fluid model can be handled similarly for the bufferless case with the assumption that for every $m = 1, \ldots, M$

$$\lim_{t\to 0} \frac{X^m(0,t)}{t} = \bar{X}^m \ a.e.$$

for some r.v. $\bar{X}^m$. Then this limit r.v. can be taken as the instantaneous fluid input rate for the calculations.

It is not always easy to find or measure the rate function for a traffic. But when the input traffic is bounded, we can find a lower bound for the rate function and hence make the admission control based on this bound. Let $\tilde{X}^m$ be one of the Nn_m sources which make up X^m let $\tilde{X}^m(t_1, t_2)$ be the amount of traffic it brings in time interval (t_1, t_2) . Assume that

$$\tilde{X}^m(t_1, t_2) \leq \pi_m(t_2 - t_1)$$

For the instantaneous rate denoted by $\tilde{X}_0^m$, we get the following from Hoeffding's Inequality:

$$\mathbb{E}[\exp\{\theta \tilde{X}_0^m\}] \leq \frac{\rho_m}{\pi_m} e^{\theta \pi_m} + \frac{\pi_m - \rho_m}{\pi_m}$$

Therefore,

$$I^{X^m}(x) = \sup_\theta\{\theta x - \log \mathbb{E}[\exp(\tilde{X}_0^m)]\} \geq \frac{x}{\pi_m} \log\left(\frac{x(\pi_m - \rho_m)}{\rho_m(\pi_m - x)}\right) - \log(\frac{\pi_m - \rho_m}{\pi_m - x})$$

We can find an on-off source for which the lower bound of the rate function is achieved. For example, choose $\tilde{X}_t^m$ to be the stationary version of the following periodic function (in the discrete time, choose time intervals small enough to make the approximation better):

$$Z_t = \begin{cases} \pi_m & 0 \leq t \leq \sigma_m/\pi_m \\ 0 & \sigma_m/\pi_m \leq t \leq \sigma_m/\rho_m \end{cases}$$

References

[1] J-Y. Le Boudec. Application of network calculus to guaranteed service networks. *IEEE Trans. Inform. Theory*, 44(3):1087–1096, 1998.

[2] C-S. Chang. On deterministic traffic regulation and service guarantees: a systematic approach by filtering. *IEEE Trans. Inform. Theory*, 44(3):1097–1110, 1998.

[3] C. Courcoubetis and R.R. Weber. Buffer overflow asymptotics for a switch handling many traffic sources. *J. Appl. Prob.*, 33:886–903, 1996.

[4] D. Y. Eun and N. Shroff. Network Decomposition in the Many-Sources Regime. *Advances in Applied Probability*, submitted, 2002.

[5] N. Likhanov and R. R. Mazumdar. Cell loss asymptotics in buffers fed with a large number of independent stationary sources. *Journal of Applied Probability*, 36(1):86–96, 1999.

[6] J. W. Roberts. *COST 242: Methods for the performance evaluation and design of broadband multiservice networks.* Elsevier N.V., 1998.

[7] A. Simonian and J. Guibert. Large deviations approximation for fluid queues fed by a large number of ON/OFF sources. *IEEE J. Sel. Areas Commun.*, 13(6):1017–1027, 1995.

[8] D. Wischik. Output of a switch, or, effective bandwidths for networks. *Queueing Systems*, 32:383–396, 1999.