DE GRUYTER OPEN Acta Univ. Sapientiae, Mathematica, **6**, 2 (2014) 146–161

DOI: 10.1515/ausm-2014-0024

Stabilizing priority fluid queueing network model

Amina Angelika Bouchentouf
Mathematics Laboratory, Djillali Liabes University of Sidi Bel Abbes, B.P. 89, Sidi Bel Abbes, Algeria
email: bouchentouf_amina@yahoo.fr

Hanane Sakhi
Department of Mathematics, Sciences and Technology University of Oran: Mohamed Boudiaf, USTOMB, B.P. 1505 EL-M'NAOUAR- Oran, Algeria
email: sakhi.hanane@yahoo.fr

Abstract. The aim of this paper is to establish the stability of fluid queueing network models under priority service discipline. To this end, we introduce a priority fluid multiclass queueing network model, composed of N stations, $N \geq 3$ and $2N$ classes (2 classes at each station); where in the system, each station may serve more than one job class with differentiated service priority, and each job may require service sequentially by more than one service station. In this paper the fluid model approach is employed in the study of the stability.

1 Introduction

Stochastic processing networks arise as models in manufacturing, telecommunications, computer systems and service industry. Common characteristics of these networks are that they have entities, such as jobs, customers or packets, that move along routes, wait in buffers, receive processing from various resources, and are subject to the effects of stochastic variability through such quantities as arrival times, processing times, and routing protocols. Networks arising in modern applications are often highly complex and heterogeneous. Typically, their analysis and control present challenging mathematical problems. One approach to these challenges is to consider approximate models.

2010 Mathematics Subject Classification: 60K25, 68M20, 90B22.
Key words and phrases: stability, fluid models, multiclass queueing networks, fluid approximation

In the last 15 years, significant progress has been made on using approximate models to understand the stability and performance of a class of stochastic processing networks called open multiclass HL queueing networks. HL stands for a non-idling service discipline that is head-of-the-line, i.e., jobs are drawn from a buffer in the order in which they arrived. Examples of such disciplines are FIFO and static priorities. First order (functional laws of large numbers) approximations called fluid models have been used to study the stability of these networks, and second order (functional central limit theorem) approximations which are diffusion models, have been used to analyze the performance of heavily congested networks.

The development of the fluid approach was inspired by the studies of some counter-examples in Kumar and Seidman [11], Rybko and Stolyar [14] and Bramson [1], etc., where the multiclass queueing networks are not stable even when the traffic intensity of each station in the network is less than one. An elegant result of the fluid model approach was proposed first in Rybko and Stolyar [14] and then generalized and refined by Dai [6], Chen [2], Dai and Meyn [8], Stolyar [15] and Bramson [1]. It states that a queueing network is stable if its corresponding fluid network model is stable. Partial converse to this result is also given in Meyn [12], Dai [7] and Puhalskii and Rybko [13]. Heng Quing Ye [10] used Kumar-Rybko-Seidman-Stolyar network for establishing the stability of fluid queueing network.

In this paper, we concentrate with the capacity of some large classes of fluid multiclass queueing networks under priority service discipline. Specifically, we establish a stability condition of some heterogenous priority fluid networks with N stations and $2\mathsf{N}$ job classes, where in the system, each station may serve more than one job class with differentiated service priority, and each job may require service sequentially by more than one service station. So, in our case, the network performance is improved even when more workloads are admitted for service. To stabilize our networks a number of stations should be added, these later act as regulators for the systems, adding these stations is not random, it depends essentially on higher and lower priority job classes (many-to-one mapping) and on the number of stations in the network. The fluid model approach is employed to proof the stability.

The outline of the paper is as follows: At first (Section 2) we describe priority fluid multiclass queueing models, and present a powerful result on the stability of such systems given by Chen and Zhang [5], after that (Section 3) we introduce modified networks and present their stability conditions (Theorems 2 and 3), and finally we conclude this paper with a short conclusion.

2 N-stations priority fluid multiclass network models

We describe N-stations priority fluid queueing network models as $(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$. Specifically, the fluid network consists of J stations (buffers) $(J = N)$ indexed by $j \in \mathcal{J} = \overline{1, N}$, serving K, $K = 2N$ fluid (customer) classes indexed by $k \in \mathcal{K} = \overline{1, 2N}$. A fluid class is served exclusively at one station, but one station may serve more than one fluid classes. $\sigma(.)$ denotes a many-to-one mapping from $\mathcal{K}$ to $\mathcal{J}$, with $\sigma(k)$ indicating the station at which a class k fluid is served. A class k fluid may flow exogenously into the network at rates λ_1 and λ_{N+1}, (≥ 0), then it is served at station $\sigma(k)$, with mean service time $m_k = 1/\mu_k$, $k = \overline{1, 2N}$ and after being served, a fraction p_{kl} of fluid turns into a class l fluid, $l \in \mathcal{K}$, and the remaining fraction, $1 - \sum_{l=1}^{K} p_{kl}$ flows out of the network. Let $C(j)$ be the set of classes that reside in station j, alternatively, we denote by a $J \times K$ matrix $C = (c_{ij})_{J \times K}$, known as the constituent matrix, where $c_{jk} = 1$ if $\sigma(k) = j$, and $c_{jk} = 0$ otherwise.

Let $Q_k(t)$ indicates the number of class k customers in the network at time t, $(Q(0) = Q_k(0))$ and $\lambda = (\lambda_k)$ two K-dimensional nonnegative vectors. $P = (p_{kl})_{K \times K}$ a stochastic matrix with spectral radius strictly less that one, $\mu = (\mu_k)$ a K-dimensional positive vector.

The vectors $Q(0)$ are referred to as initial fluid level vector, λ to the exogenous inflow rate vector, μ to the processing rate vector, matrix P is referred to as flow transfer matrix.

When station $\sigma(k)$ devotes its full capacity to serving class k fluid (assuming that it is available to be served), it generates an outflow of class k fluid at rate $\mu_k > 0$, $k \in \mathcal{K}$. Among classes, fluid follows a priority service discipline, which is again described by a one-to-one mapping π from $\{1, ..., K\}$ onto itself. Specifically, a class k has priority over a class l if $\pi(k) < \pi(l)$ and $\sigma(l) = \sigma(k)$, then class k job can not be served at station $\sigma(k)$ unless there is no class l job.

So, our multiclass fluid network consists of N stations and $2N$ job classes. Assume that the arrival process of class k, $k = \overline{1, 2N}$, customers arrive to the system following a Poisson process with arrival rates $\lambda_1 \geq 0$ and $\lambda_{N+1} \geq 0$, the service time for each class k customer is exponentially distributed with mean service time $m_k > 0$. We also assume that all the inter-arrival times and service times are independent.

To describe the dynamics of the fluid network, we introduce the K-dimensional fluid level process $\overline{Q} = \{\overline{Q}(t), \ t \geq 0\}$, whose kth component $\overline{Q}_k(t)$ denotes the fluid level of class k at time t; the K-dimensional time allocation process $\overline{T} = \{\overline{T}(t), t \geq 0\}$, whose kth component $\overline{T}_k(t)$ denotes the total amount of time that station $\sigma(k)$ has devoted to serving class k fluid during the time interval $[0,t]$, and the K-dimensional unused capacity process $\overline{Y} = \{\overline{Y}(t), t \geq 0\}$, whose kth component $\overline{Y}_k(t)$ denotes the (cumulative) unused capacity of station $\sigma(k)$ during the time interval $[0,t]$ after serving all classes at station $\sigma(k)$ which have a priority no less than class k. We denote by D the K-dimensional diagonal matrix whose kth element is μ_k, and e is a K-dimensional vector with all components being one. Let

$$H_k = \{l : \sigma(l) = \sigma(k), \pi(l) \leq \pi(k)\}$$

be the set of indices for all classes that are served at the same station as class k and have a priority no less than that of class k. Note that $k \in H_k$ by definition. Then the dynamics of the fluid network model can be described as follows.

$$\overline{Q}(t) = \overline{Q}(0) + \lambda t - (I - P')D\overline{T}(t) \geq 0, \tag{1}$$

$$\overline{T}(\cdot) \text{ are nondecreasing with } \overline{T}(0) = 0, \tag{2}$$

$$\overline{Y}_k(t) = t - \sum_{l \in H_k} \overline{T}_l(t) \text{ are nondecreasing, } k \in \mathcal{K}, \tag{3}$$

$$\int_0^\infty \overline{Q}_k(t) d\overline{Y}_k(t) = 0, \ k \in \mathcal{K}. \tag{4}$$

Let

$$Q_k(t) = Q_k(0) + \lambda_k t + \sum_{l=1}^{K} p_{lk}\mu_l T_l(t) - \mu_k T_k(t) \geq 0, k = 1, \ldots, K, \tag{5}$$

be the kth coordinate of the flow balance relation (1).

The equation (1) is the equivalent relation between the time allocation process $T(\cdot)$ and the unused capacity process $Y(.)$. The relation (4) means that at any time t, there could be some positive remaining capacity (rate) for serving those classes at station $\sigma(k)$ having a strictly lower priority than class k, only when the fluid levels of all classes in H_k (having a priority no less than k) are zero.

A pair $(\overline{Q}, \overline{T})$ (or equivalently $(\overline{Q}, \overline{Y})$) is said to be a fluid solution if they jointly satisfy (1)-(4). For convenience, we also call $\overline{Q}$ a fluid solution if there is a $\overline{T}$ such that the pair $(\overline{Q}, \overline{T})$ is a fluid solution. The fluid network $(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$ is said to be stable if there is a time $\tau \geq 0$ such that $\overline{Q}(\tau + \cdot) \equiv 0$ for any fluid solution $\overline{Q}$ with $\|\overline{Q}(0)\| = 1$; and it is said to weakly stable if $\overline{Q}(\cdot) = 0$ for any fluid solution $\overline{Q}$ with $\overline{Q}(0) = 0$. The processes $\overline{Q}$, $\overline{Y}$, and $\overline{T}$ are Lipschitz continuous, and hence are differentiable almost everywhere on $[0, \infty)$, this well-known property will be used in this paper.

It is well-known that the queue length process $Q(t)$ is a continuous time Markov chain under the Poisson arrival and exponential service assumptions. We say that the network $(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$ is stable if the Markov chain $Q(t)$ is positive recurrent. It is well-know that the Markov chain $Q(t)$ is positive recurrent only if the traffic intensity for each station is less than one, i.e., $\rho_j < 1$ (ρ_j is the jth component of ρ; a traffic intensity for station j) for all $j \in \mathcal{J}$, or in short, $\rho < e$, where e is a J-dimensional vector with all components being ones.

The expected stationary total queue length $\overline{Q}$ is defined as

$$\overline{Q} = \lim_{t \to \infty} \mathbb{E}\left[\sum_{k \in \mathcal{K}} Q_k(t)\right].$$

The queue length $\overline{Q}(t)$ is a finite if and only if the queue length process Q is positive recurrent.

Chen and Zhang [5] gave a very important result on the stability of priority fluid queueing systems, authors established the sufficient condition for the stability based on the existence of a linear Lyapunov function, this later (sufficient condition) gave the the necessary and sufficient condition for the stability. Their result is presented in the Theorem 1, in order to state it we need some additional assumptions:

Let

$$h(k) = \begin{cases} \arg\max\{\pi(l) : l \in H_k^+\} & \text{if } H_k^+ \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

with $H_k^+ = H_k \backslash \{k\}$, in words; if k is not the highest priority class at station $\sigma(k)$, then $h(k)$ is the index for the class which has the next higher priority than class k at station $\sigma(k)$, otherwise $h(k) = 0$.

$$\theta = \lambda - (I - P')\mu_H^0, \tag{7}$$

where $\mu_H^0 = De_H^0$, $(e_H^0 = (e_1^0, ..., e_K^0)')$ is a K-dimensional vector with $e_k^0 = 1$ if $H_k^+ = \emptyset$ and $e_k^0 = 0$ otherwise.

$$R = (I - P')D(I - B), \tag{8}$$

where $B = (b_{lk})$ is the $K \times K$ matrix with $b_{lk} = 1$ if $k = h(l)$, and $b_{lk} = 0$ otherwise, $(l, k = 1, ...K)$.

And let

$$\rho = CD^{-1}(I - P')^{-1}\lambda \quad (9)$$

be the traffic intensity of the queueing network.

Theorem 1 [5] *Consider a fluid network* (λ, μ, P, C) *under priority service discipline* π. *Let vector* θ *and matrix* R *be as defined in (7),(8) respectively. Assume that* $\rho < e$. *Then the fluid network is stable if there exist a K-dimensional vector* $h \geq 0$ *such that for any given partition* a *and* b *of* $\mathcal{K}$ *satisfying if class* $l \in a$, *then each class k with*

$$\sigma(k) = \sigma(l) \ \textit{and} \ \pi(k) > \pi(l) \ \textit{is also in} \ a, \quad (10)$$

we have

$$h'_a(\theta_a + R_{ab}x_b) < 0 \quad (11)$$

for $x_b \in \mathcal{S}_b := \{u \geq 0 : \theta_b + R_b u = 0 \ \textit{and} \ u \leq e\}$ *when* $b \neq \emptyset$, *and* $x_b = 0$ *when* $b = \emptyset$. *The inequality (11) is omitted to hold by default when* $S_b = \emptyset$.

Set a includes all classes which have zero unused capacity rate and set b includes all classes which have a positive unused capacity rate at time t.

3 Main result

In this paper, we present two theorems, we provide the proof of the first theorem, while the proof of the second one is omitted since it is similar to the former one.

3.1 Stabilizing N-stations priority fluid queueing network with some additional stations

Our multiclass queueing network consists of N stations and $2N$ job classes. Assume that the arrival process of class k; $k = \overline{1, 2N}$ customers arrive to the system following a Poisson process with arrival rates λ_1 and λ_{N+1} (≥ 0) to station 1 and $N + 1$ respectively, the service time for each class k customer is exponentially distributed with mean service time $m_k > 0$. We also assume that all the inter-arrival times and service times are independent.

Suppose that each even class at station $i = \overline{1, N}$ has higher priority.

We modify our network such that if it is composed of an even number of stations we add N additional ones otherwise we add $(N-1)$, the explanation of this choice will be given in the rest of the paper, the modified network is illustrated in Figures 1 and 2; the additional stations are named station $N+1$,...,station $2N$, (N: even) (resp. station $N+1$,...,station $2N-1$ (N: odd)).

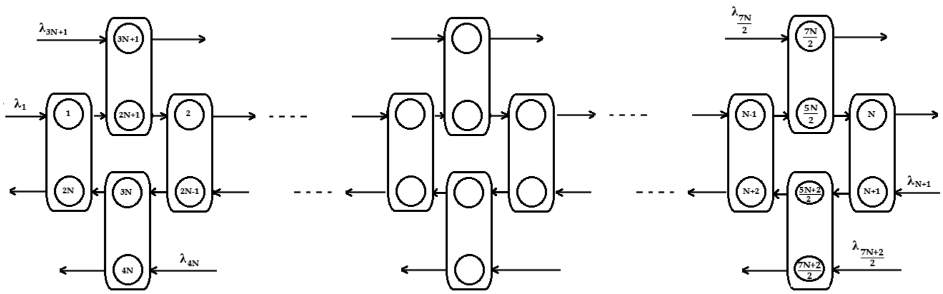

Figure 1: 2N-stations priority fluid queueing network

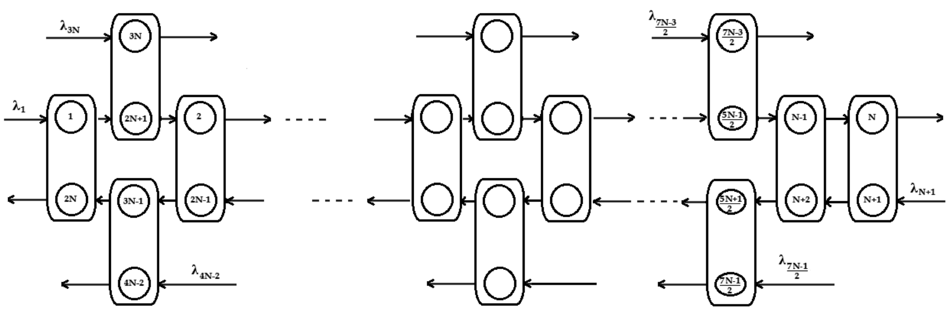

Figure 2: 2N-1-stations priority fluid queueing network

Theorem 2 *Suppose* $\rho < e$, *equation (11) not satisfied.*

If

$$\lambda_{k_1} > (1 - m_{k_1'}/m_{k_1''})/m_{k_1} \tag{12}$$

then the queue length process $Q(\cdot)$ *is positive recurrent.*

λ_{k_1} *(resp.* m_{k_1}*) is the exogenous arrival rate (resp. the mean service time) of higher priority fluid class of additional stations* $i = \overline{N+1, 2N}$, *(N: even) (resp.* $i = \overline{N+1, 2N-1}$, *(N: odd)), such that* $k_1 = \overline{3N+1, 4N}, (N : \text{even})$ *(resp.* $k_1 = \overline{3N, 4N-2}$, $(N : \text{odd})$*).* $m_{k_1'}$ *is the mean service time of lower* *priority fluid class of additional stations* $i = \overline{N+1, 2N}$, *(N: even) (resp.* $i = \overline{N+1, 2N-1}$, *(N: odd)).* $m_{k_1''}$ *is the mean service time of higher priority fluid class of the original network.*

With

$$m_{k_1'} = \begin{cases} m_{k_1-N}, & k_1 = \overline{3N+1, 4N}, \quad \text{(N: even)}, \\ m_{k_1-(N-1)}, & k_1 = \overline{3N, 4N-2}, \quad \text{(N: odd)}. \end{cases}$$

$$m_{k_1''} = \begin{cases} m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N}{2}}, \ j_1 = \overline{1, \frac{N}{2}} \\ & \qquad\qquad\qquad\qquad \text{(N: even)}, \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+2}{2}, k_1 - N}, \ j_1 = \overline{2, N} \\ m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N-1}{2}}, \ j_1 = \overline{1, \frac{N-1}{2}} \\ & \qquad\qquad\qquad\qquad \text{(N: odd)}. \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+1}{2}, k_1-(N-1)}, \ j_1 = \overline{4, N+1} \end{cases}$$

Where for each k_1' *it corresponds* k_1 *and* j_1, *(*j_1 *is an even number).*

Via this theorem, we show that when the arrival rates of some job classes is reduced, the performance of the queue will worse.

Proof. In Chen and Yao [3] and Dai [7], it was shown that to prove the stability of a queueing model, it is sufficient to study the stability of its corresponding fluid queueing model, our prove is based on this result. To understand better the phenomenon, let us examine the dynamics of the original network with no initial job. When the higher priority job classes are being served, the lower priority ones are in standby, waiting for service, (class 1 jobs can not move to class 2 and 2 cannot move to 3,... for further services, and vice versa). So, these classes will never be served at the same time and in effect form virtual stations (Dai and Vande Vate [9]). Therefore, the total nominal traffic intensity for these classes together, i.e., the virtual stations, should not exceed one for the network to be stable. The similar argument also yields that the network is unstable when the nominal traffic intensity for the virtual stations exceed one, i.e., the condition (11) is not satisfied. Now consider the modified network. The additional classes act as regulators that regulate the traffics in the network so as to stabilize the network. When the workloads of classes k_1 (k_1; defined in the theorem) are light, much service capacity of the additional stations are left to classes k_1' (k_1'; defined in the theorem) and hence these later do not hold back the traffics to avoid building up of job queues at priority classes of the original network. Thus, the virtual stations effect prevails and the network is still unstable when the condition (11) is not satisfied. However, when the workloads of classes k_1 are heavy enough such that the condition (12) holds, the service for lower priority classes at additional stations is in effect slowed down and the traffics in the original network are held back (these classes will not mutually block their services). Finally, the virtual station effect is avoided and the modified network is thus stabilized.

The dynamics of the our modified fluid network model can be described as follows.

$$\overline{Q}_{k_1}(t) = \overline{Q}_{k_1}(0) + \lambda_{k_1} t - \mu_{k_1} \overline{T}_{k_1}(t) \geq 0, \tag{13}$$

$k_1 = 1, N+1, \overline{3N+1, 4N}, (N : even), (resp.\ k_1 = 1, N+1, \overline{3N, 4N-2})$
$(N : \mathrm{odd})$,

$$\overline{Q}_k(t) = \overline{Q}_k(0) + \mu_l \overline{T}_l(t) - \mu_k \overline{T}_k(t) \geq 0, \tag{14}$$

(k, l) = two successive job classes, such that the k^{th} class is the arriving l^{th} class

$$\overline{T}_k(\cdot) \text{ are nondecreasing with } \overline{T}_k(0) = 0, \tag{15}$$

$$k = \overline{1, 4N}, (N : even) (\text{resp. } k = \overline{1, 4N-2}, (N : \mathrm{odd}))$$

$$\begin{cases} \overline{Y}_{k_1}(t) = t - \overline{T}_{k_1}(t), \\ \\ \overline{Y}_{k_1''}(t) = t - \overline{T}_{k_1''}(t), \end{cases} \quad \text{are nondecreasing,} \tag{16}$$

$$\overline{Y}_k(t) = t - \overline{T}_l(t) - \overline{T}_k(t) \text{ are nondecreasing,} \tag{17}$$

(k, l) = (lower priority job class, higher priority job class) at station i, $i = \overline{1, 2N}$, $(N : even)$ (resp. $i = \overline{1, 2N-2}$, $(N : \mathrm{odd})$),

$$\int_0^\infty \overline{Q}_k(t) d\overline{Y}_k(t) = 0, \;\; k = \overline{1, 4N}(N : even) \;\; (\text{resp. } k = \overline{1, 4N-2}(N : \mathrm{odd})). \tag{18}$$

The stability study of the modified fluid network will be done in three steps.

1. First step. We prove that there exists a time $\tau_1 \geq 0$ such that

$$\overline{Q}_{k_1}(t) = 0, \quad \text{for any } t \geq \tau_1, \tag{19}$$

with $k_1 = \overline{3N+1, 4N}$, (N: even), (resp. $k_1 = \overline{3N, 4N-2}$, (N: odd)).

If $\dot{\overline{Q}}_{k_1}(t) > 0$, then we have by equation (18)

$$\dot{\overline{Y}}_{k_1}(t) = 0, \tag{20}$$

then by conditions (16) and (20)

$$\dot{\overline{T}}_{k_1}(t) = 1, \tag{21}$$

then by (13) and (21), we get

$$\dot{\overline{Q}}_{k_1}(t) = \lambda_{k_1} - \mu_{k_1}. \tag{22}$$

Note that the condition $\rho < e$ implies $\lambda_{k_1} < \mu_{k_1}$. Let $\tau_1^{(l)} = \dot{\overline{Q}}_{k_1}(0)/(\mu_{k_1} - \lambda_{k_1})$, $l = \overline{1, \frac{N}{2}}$, (N: even) (resp. $l = \overline{1, \frac{N-1}{2}}$, (N: odd)). Then, we have

$$\overline{Q}_{k_1}(t) = 0 \text{ for any } t \geq \tau_1^{(l)}. \tag{23}$$

Letting $\tau_1 = \max(1/\mu_{k_1} - \lambda_{k_1})$, we have that $\tau_1 \geq \max(\tau_1^{(l)})$, (each l corresponds to k_1) under the assumption $\|\overline{Q}(0)\| = 1$. Now, the conclusion (23) leads to the claim (19).

2. ***Second step.*** We prove that there exists a time $\tau_2 \geq \tau_1$ such that

$$\overline{Q}_{k_1''}(t) = 0, \text{for any } t \geq \tau_2, \tag{24}$$

where k_1'' is the higher priority job class at station i, $i = \overline{1, N}$.

$$k_1'' = \begin{cases} k_1' - (2N - j_1), & k_1' = \overline{2N+1, \frac{5N}{2}}, \ j_1 = \overline{1, \frac{N}{2}} \\ & \qquad\qquad\qquad\qquad \text{(N: even)}, \\ (k_1 - k_1') + j_1, & k_1' = \overline{\frac{5N+2}{2}, k_1 - N}, \ j_1 = \overline{2, N} \\ k_1' - (2N - j_1), & k_1' = \overline{2N+1, \frac{5N-1}{2}}, \ j_1 = \overline{1, \frac{N-1}{2}} \\ & \qquad\qquad\qquad\qquad \text{(N: odd)}. \\ (k_1 - k_1') + j_1, & k_1' = \overline{\frac{5N+1}{2}, k_1 - (N-1)}, \ j_1 = \overline{4, N+1} \end{cases}$$

Under the condition (19), we have $\dot{\overline{Q}}_{k_1}(t) = 0$, and then $\dot{\overline{T}}_{k_1}(t) = \lambda_{k_1} m_{k_1}$, $k_1 = \overline{3N+1, 4N}$, (N: even) (resp. $k_1 = \overline{3N, 4N-2}$, (N: odd)), for all time $t \geq \tau_1$. Combined with (17), this gives rise to

$$\dot{\overline{Y}}_{k_1'}(t) = t - \dot{\overline{T}}_{k_1'}(t) - \dot{\overline{T}}_{k_1}(t) \geq 0,$$

k_1' are classes of lower priority at additional stations,

$$k_1' = \begin{cases} k_1 - N, & k_1 = \overline{3N+1, 4N}, \ \text{(N: even)}, \\ k_1 - (N-1), & k_1 = \overline{3N, 4N-2}, \ \text{(N: odd)}. \end{cases}$$

and

$$\dot{\overline{T}}_{k_1'}(t) \leq 1 - \dot{\overline{T}}_{k_1}(t) = 1 - \lambda_{k_1} m_{k_1}, \ \text{for any} \ t \geq \tau_1. \tag{25}$$

Then,

$\dot{\overline{Q}}_{k_1''}(t) = \mu_{k_1'} \dot{\overline{T}}_{k_1'}(t) - \mu_{k_1''} \dot{\overline{T}}_{k_1''}(t) \leq \mu_{k_1'}(1 - \lambda_{k_1} m_{k_1}) - \mu_{k_1''} < 0$, where for each k_1' it corresponds k_1'' for any $t \geq \tau_2$, where the last inequality is implied by the assumption that

$$\lambda_{k_1} > (1 - m_{k_1'}/m_{k_1''})/m_{k_1}.$$

Let $\tau_2^{(l)} = \frac{\dot{\overline{Q}}_{k_1''}(\tau_1)}{\mu_{k_1''} - \mu_{k_1'}(1-\lambda_{k_1} m_{k_1})}$, $l = \overline{1, N}$ (N: even), (resp. $l = \overline{1, N-1}$ (N: odd)). Then, we have

$$\overline{Q}_{k_1''}(t) = 0 \ \text{for any} \ t \geq \tau_2^{(l)}. \tag{26}$$

Let

$$\tau_2 = \max\left(\frac{1 + \Theta\tau_1}{\mu_{k_1''} - \mu_{k_1'}(1 - \lambda_{k_1} m_{k_1})}\right)$$

with Θ being the Lipschitz constant for the fluid level process $\overline{Q}(t)$. Then we have that $\tau_2 \geq \max(\tau_2^{(l)})$. Now, the conclusion (26) implies the claim (24).

Before to pass to the last step, we prove separately that $\overline{Q}_{N+1}(t) = 0$ for any $t \geq \tau_2$, "the case of network with even number of stations".
If $\dot{\overline{Q}}_{N+1}(t) = 0$, this implies $\dot{\overline{Y}}_{N+1}(t) = 0$, which in turn implies that $\dot{\overline{T}}_{N+1}(t) = 1$, then we have $\dot{\overline{Q}}_{N+1}(t) = \lambda_{N+1} - \mu_{N+1}$, with $\lambda_{N+1} < \mu_{N+1}$ (since $\rho < e$). So there exists $\tau_2' = \dot{\overline{Q}}_{N+1}(0)/\mu_{N+1} - \lambda_{N+1}$, such that $\dot{\overline{Q}}_{N+1}(t) = 0$ for any $t \geq \tau_2$.

Third step. We prove that there exists a time $\tau \geq \tau_2$ (≥ 0) such that

$$\overline{Q}_l(t) = 0, \ \text{for} \ t \geq \tau, \tag{27}$$

l represents job classes of lower priority at station $i = \overline{1, 2N}$ (N: even) (resp. $i = \overline{1, 2N-1}$ (N: odd), which together with equations (19) and (24) implies

$$\overline{Q}(t) = 0 \ \text{for} \ t \geq \tau.$$

Let

$$\overline{W}_i(t) = (\lambda_1 m_{l_1} + \lambda_{N+1} m_{l_2})t - \sum_{k:\sigma(k)=i} \overline{T}_k(t),\ i = \overline{1,N},$$

with $l_1 = \overline{1,N}$ and $l_2 = \overline{N=1,2N}$ job classes at the same station in the original network.

$\overline{W}_{i'}(t)$=

$$\begin{cases} \lambda_1 m_{k_1'} t - \overline{T}_{k_1'}(t), & k_1' = \overline{2N+1, \frac{5N}{2}}, \text{N: even}, \\ & (resp.\ k_1' = \overline{2N+1, 2N+\frac{5N-1}{2}}, \text{N: odd}) \\ \lambda_{N+1} m_{k_1'} t - \overline{T}_{k_1'}(t), & k_1' = \overline{\frac{5N+2}{2}, 3N}, \text{N: even}, \\ & (resp.\ k_1' = \overline{\frac{5N+1}{2}, 3N-1}, \text{N: odd}) \end{cases}$$

for $\tau \geq \tau_2$. Here $\overline{W}(t)$ can be explained as the immediate workload in the system at time t. Define

$f_i(t) = k_1' \overline{W}_i(t)$, with k_1' a lower priority job class in the additional stations.
$f_{i'}(t) = k_1'' \overline{W}_{i'}(t)$, with k_1'' a higher priority job class in the original network.
For each i (resp. i') it corresponds to k_1' (resp. k_1'').

Then, it is direct to verify that, for $t \geq \tau_2$,

$\dot{f}_i(t) < 0$ if $\dot{\overline{Q}}_i(t) > 0$, for $i = \overline{1,4N}$, (N: even, (resp. $i = \overline{1,4N-2}$, (N: odd))

And

$$f_1(t) \leq f_N(t) \text{ if } \overline{Q}_1(t) = 0,$$

$$f_i(t) \leq f_{i-1}(t) \text{ if } \overline{Q}_i(t) = 0,\ i = \overline{2,3N}, (\text{N: even}) \quad (\text{resp. } i = \overline{2,3N-1}, (\text{N: odd}))$$

$$f_j(t) \leq f_i(t) \text{ if } \sum_{j \neq i} \overline{Q}_j(t) = 0,\ j = \overline{1,3N}, (\text{N: even}) \quad (\text{resp. } j = \overline{1,3N-1}, (\text{N: odd}))$$

$$f_N(t) \leq f_N(t) \text{ if } \overline{Q}_{3N}(t) = 0, (\text{N: even}) \ (\text{resp. } \overline{Q}_{3N-1}(t) = 0, (\text{N: odd}))$$

Now applying the piecewise linear Lyapunov function approach for the multiclass fluid network model described in Theorem 3.1 of Chen and Ye [4], we obtain the conclusion (27). □

3.2 Stabilizing N-stations priority fluid queueing networks with N additional stations

Our N-stations multiclass queueing network is the same is above. Suppose in this case that the higher priority is devoted to classes $N, \overline{N+2, 2N}$.

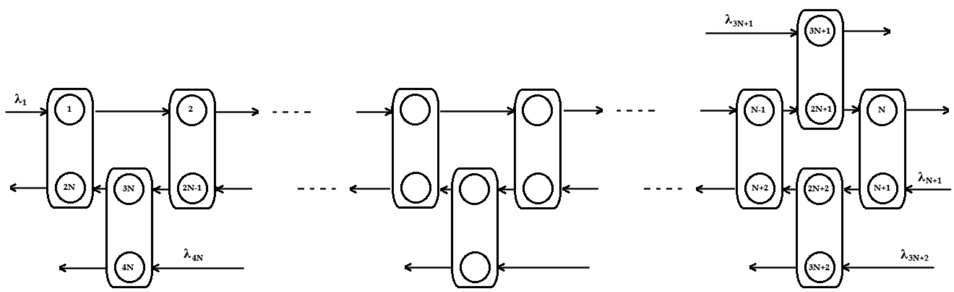

Figure 3: 2N-stations priority fluid queueing network

We modify our network by adding N stations, (see Figure 3), compared to the original network, there are N additional stations, namely the station $N+1, \ldots,$ station 2N, such that, $3N+1$ job class has high priority at station $N+1$, and $\overline{3N+2,\ 4N}$ job classes have higher priority at stations $\overline{N+2, 2N}$. Now, let us introduce the second main result.

Theorem 3 *Suppose* $\rho < e$ *holds, equation (11) not satisfied.*
If

$$\lambda_{3N+1} > (1 - m_{2N+1}/m_N)/m_{3N+1}, \ \ \lambda_{k_2} > (1 - m_{k_2'}/m_{k_2''})/m_{k_2}, \tag{28}$$

$k_2'' = \overline{N, 2N}$, $k_2' = \overline{2N+1, 3N}$, $k_2 = \overline{3N+1, 4N}$, *where for each* k_2 *it corresponds to* k_2' *and* k_2''.

Then the queue length process $Q(\cdot)$ *is positive recurrent.*

In this case, when the higher priority classes are being served, the lower priority ones cannot be served, (class 1 cannot move to class 2 and 2 cannot move to 3,..., for further service, and vice versa. So, these later form a virtual stations. Therefore, these later, should not exceed one for the network to be stable. Now, let us consider the modified network. The additional classes 2N+1 and $\overline{3N+1, 4N}$ act as regulators that regulate the traffics. When the workloads of classes $3N+1$ and $\overline{3N+2, 4N}$ are light, much service capacity of stations

$N+1,\dots,2N$ are left to classes $\overline{2N+1,3N}$ respectively and hence these later do not hold back the traffics to avoid building up of job queues at higher priority classes of the original network. Thus, the virtual stations effect prevails and the network is still unstable. However, when the workloads of classes $\overline{3N+1,4N}$ are heavy enough such that the condition (28) holds, the service for lower priority classes $\overline{2N+1,3N}$ is in effect slowed down and the traffics to the higher priority classes N and $\overline{N+2,2N}$ are held back. Finally, the virtual stations effect is avoided and the modified network is thus stabilized.

Then, following the same steps given in theorem 2, it is not difficult to prove that there exists a time $\tau_1 \geq 0$ such that

$$\overline{Q}_{k_2}(t) = 0, \;\; k_2 = \overline{3N+1,4N}, \;\; \text{for any} \;\; t \geq \tau_1. \tag{29}$$

after that, we prove that there exists a time $\tau_2 \geq \tau_1$ such that

$$\overline{Q}_N(t) = \overline{Q}_{k_2''}(t) = 0, \;\; k_2'' = \overline{N+2,2N} \;\; \text{for any} \;\; t \geq \tau_2. \tag{30}$$

and finally, we prove that there exists a time $\tau \geq \tau_2 (\geq 0)$ such that

$$\overline{Q}_{k_4'}(t) = 0 \quad k_4' = \text{lower priority job class at stations } i = \overline{1,2N}, \;\; \text{for} \;\; t \geq \tau. \tag{31}$$

4 Conclusion

Multiclass queueing networks are effective tools for modelling many industrial settings. One setting for which the model is particularly attractive is the production flow within semiconductor manufacturing facilities.

In this paper we have studied the stabilization of N-stations queueing networks using its corresponding fluid network. The resulting model, fluid queueing networks with additional stations depending on the service priority and on the number of stations in the network are formally presented in Section 3. Beyond the presentation of our modified network models "fluid networks with additional stations", the primary concern of the paper is the stability of such networks. Nevertheless, stability of the artificial fluid model implies stability of the original network (see Theorems 2 and 3).

References

[1] M. Bramson, Stability of two families of queueing networks and a discussion of fluid limits, *Queueing Syst.*, **23** (1998), 7–31.

[2] H. Chen, Fluid approximations and stability of multiclass queueing networks: Workconserving discipline, *Ann. Appl. Probab.*, **5** (1995), 637–655.

[3] H. Chen, D. D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization*, Springer-Verlag New York, Inc. (2001).

[4] H. Chen, H. Q. Ye, Piecewise linear Lyapunov function for the stability of priority multiclass queueing networks, *IEEE Trans. Automat. Control*, **47** (4) (2002), 564–575.

[5] H. Chen, H. Zhang, Stability of multiclass queueing networks under priority service disciplines, *Oper. Res.*, **48** (2000), 26–37.

[6] J. G. Dai, On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid models, *Ann. Appl. Probab.*, **5** (1995), 49–77.

[7] J. G. Dai, A fluid-limit model criterion for instability of multiclass queueing networks, *Ann. Appl. Probab.*, **6** (1996), 751–757.

[8] J. G. Dai, S. P. Meyn, Stability and convergence of moments for multiclass queueing networks via fluid models, *IEEE Trans. Automat. Control*, **40** (1995), 1899–1904.

[9] J. G. Dai, J. H. Vande Vate, Global Stability of Two-Station Queueing Networks. *Proceedings of Workshop on Stochastic Networks: Stability and Rare Events, Editors: Paul Glasserman, Karl Sigman and David Yao, Springer-Verlag, Columbia University, New York.*, (1996), 1–26.

[10] H. Q. Ye, A paradox for admission control of multiclass queueing network with differentiated service, *J. Appl. Probab.*, **44** (2) (2007), 321–331.

[11] P. R. Kumar, T. I. Seidman, Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems, *IEEE Trans. Automat. Control*, **35** (1990), 289–298.

[12] S. Meyn, Transience of multiclass queueing networks via fluid limit models, *Ann. Appl. Probab.*, **5** (1995), 946–957.

[13] A. Puhalskii, A. N. Rybko, Non-ergodicity of queueing networks under nonstability of their fluid models, *Probl. Inf. Transm.*, **36** (1) (2000), 26–48.

[14] A. N. Rybko, A. L. Stolyar, Ergodicity of stochastic processed describing the operations of open queueing networks, *Problemy Peredachi Informatsii*, **28** (1992), 2–26.

[15] A. L. Stolyar, On the stability of multiclass queueing network: a relaxed sufficient condition via limiting fluid processes, *Markov Process. Related Fields*, **1** (4) (1995), 491–512.

Received: 15 November 2013