INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Smoothing Effect of the Superposition of Homogeneous Sources in Tandem Networks

Arie Hordijk — Zhen Liu — Don Towsley

N° 3839

Décembre 1999

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3839--FR+ENG

Smoothing Effect of the Superposition of Homogeneous Sources in Tandem Networks

Arie Hordijk , Zhen Liu , Don Towsley

Thème 1 — Réseaux et systèmes

Projet Mistral

Rapport de recherche n° 3839 — Décembre 1999 — 20 pages

Abstract: We analyze the smoothing effect of superposing homogeneous sources in a network. We consider a tandem queueing network representing the nodes that customers generated by these sources pass through. The servers in the tandem queues have different time varying service rates. In between the tandem queues there are propagation delays. We show that for arbitrary arrival and service processes which are mutually independent, the sum of unfinished works in the tandem queues is monotone in the number of homogeneous sources in the increasing convex order sense, provided the total intensity of the foreground traffic is constant. The results hold for both fluid and discrete traffic models.

Key-words: homogeneous sources, traffic superposition, stochastic comparison, multiplexing gain.

Unité de recherche INRIA Sophia Antipolis

2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)

Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Effet de lissage de la superposition des sources homogènes dans les réseaux tandem

Résumé : Nous analysons l'effet de lissage de la superposition des sources homogènes dans un réseau. Nous considérons un réseau tandem représentant les nœuds que traversent les trafics générés par ces sources. Les serveurs dans les files d'attente en tandem ont les taux de service différents et qui varient en temps. Entre deux files en tandem il y a un délai de propagation. Nous montrons que pour les processus des arrivées et des services arbitraires qui sont mutuellement indépendants, la somme des charges dans les files en tandem est monotone en nombre des sources homogènes dans le sens de l'ordre convexe croissant, pourvu que l'intensité totale du traffic est constante. Ce résultat est valide pour les modèles de trafic aussi bien fluid que discret.

Mots-clés : sources homogènes, superposition du trafic, comparaison stochastique, gain de multiplexage.

1 Introduction

The aggregation of traffic sources has been receiving increasing attention in the performance analysis of communication networks. The smoothing effect of superposing homogeneous sources in the framework of single-queue systems has been reported in various papers. Consider a queueing system with buffer size B and constant service rate C. When the queue is fed by a large number of sources, say n, Botvich and Duffield [5] showed that, under mild statistical assumptions, the loss probability decays exponentially in n when $B = nb$ and $C = nc$. Various results on this decay rate function $I(b)$ exist under more restrictive assumptions on the traffic sources and on b (large or small), see for example Weiss [14], Elwalid et al. [8], Botvich and Duffield [5] and Courcoubetis and Weber [6]. More recent results were obtained by Dumas and Simonian [7] for on-off sources where on-durations have a heavy tail distribution, and by Mandjes [11] for semi-Markov fluid sources.

Other analyses have been carried out on the qualitative comparison of the queue length and unfinished work. Koole and Liu [9] considered a queueing model fed by Markovian traffic sources where the background traffic is modeled by a Markov Arrival Process (MAP), and foreground traffic is modeled by $N \geq 1$ homogeneous on-off sources whose on and off durations are exponentially distributed. They showed that the queue length in the infinite-capacity buffer system (respectively the number of losses in finite-capacity buffer system) is larger in the increasing convex order sense (respectively the strong stochastic order sense) than the queue length (respectively number of losses) of the queueing system with the same background traffic and MN homogeneous on-off sources of the same total intensity as the foreground traffic, where M is an arbitrary integer. Consequently, when the foreground traffic consists of multiple homogeneous on-off sources, the queue length (respectively the loss process) is upper bounded by a similar system fed by a single on-off source and lower bounded by one fed by a Poisson source, where the comparison is in the increasing convex sense (respectively the strong stochastic sense). They also compared $N \geq 1$ homogeneous arbitrary two-state Markov Modulated Poisson Process (MMPP) sources and proved the monotonicity of the queue length in the transition rates and its convexity in the arrival rates. These results were generalized in [10] to the semi-Markovian case, where the on durations follow decreasing-failure-rate distributions It was shown that the queue length increases in the increasing convex order sense when the vector of arrival rates of the on-off sources increases in the majorization sense. As a consequence, the queue length is monotone in the number of homogeneous on-off sources in the increasing convex order sense, provided the total intensity of the foreground traffic is constant.

Parallel to these works, Bäuerle [3, 4] obtained similar results for a general fluid model. More precisely, Bäuerle showed that for a fluid single-server queueing system, the stationary workload is smaller in the increasing convex ordering sense with N homogeneous fluid sources, each with rate process $A(t)/N$, than with M homogeneous fluid sources, each with rate process $A(t)/M$, where $M < N$. When the fluid queueing system has finite-capacity buffer, the loss is smaller in the increasing convex ordering sense with N homogeneous fluid sources than with M homogeneous fluid sources, where $M < N$.

In this paper, we analyze the smoothing effect of superposing homogeneous sources generating customers that pass through a tandem queueing network. The servers in the tandem queues have different time varying service rates and customers incur fixed propagation delays when traversing from queue to queue. We show that for arbitrary arrival and service processes which are mutually independent, the sum of workloads in the tandem queueing system is monotone in the number of homogeneous on-off sources in the increasing convex order sense, provided the total intensity of the foreground traffic is constant. The results hold for both fluid and discrete traffic models.

The paper is organized as follows. In Section 2 we present some notions and preliminary results of stochastic orders and majorization. In Section 3 we analyze a queueing system with fluid traffic sources and derive comparison results of workload. In Section 4 we point out extensions of the results to the discrete traffic models. Concluding remarks are provided in Section 5.

2 Preliminaries on Majorization and Stochastic Orders

Let $\boldsymbol{x}, \boldsymbol{y} \in I\!R^n$, be two n-dimensional real-valued vectors. We introduce the notation $x_{[k]}$ to denote the *k-th largest element* in vector $\boldsymbol{x}$ and define the following ordering (see [12]).

Vector $\boldsymbol{y}$ is said to majorize vector $\boldsymbol{x}$ (written $\boldsymbol{x} \prec \boldsymbol{y}$) if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1, \cdots, n-1,$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad . \tag{1}$$

A weaker ordering can be defined by replacing the equality in (1) by an inequality. This implies, $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$, $k = 1, \cdots, n$. In this case, vector $\boldsymbol{y}$ is said to *weakly submajorize* vector $\boldsymbol{x}$ (or, weakly majorize $\boldsymbol{x}$ from below) written $\boldsymbol{x} \prec_w \boldsymbol{y}$.

Various properties related to majorization and weak majorization can be found in Marshall and Olkin [12]. We shall use in particular the characterization:

Proposition 1 (Proposition 4.B.2, page 109, [12]) *The relation $\boldsymbol{x} \prec_w \boldsymbol{y}$ holds if and only if for all continuous increasing and convex function $g : I\!R \to I\!R$, $\sum_{i=1}^{n} g(x_i) \leq \sum_{i=1}^{n} g(y_i)$.*

As consequences we can easily show

Corollary 2 *Let m and n be two integers, $\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in I\!R^n$ and $\boldsymbol{y} = (y_1, y_2, \cdots, y_n) \in I\!R^n$, $\boldsymbol{u} = (u_1, u_2, \cdots, u_m) \in I\!R^m$ and $\boldsymbol{v} = (v_1, v_2, \cdots, v_m) \in I\!R^m$. Then*

(a) $\boldsymbol{x} \prec_w \boldsymbol{y}$ implies $(f(x_1), f(x_2), \cdots, f(x_n)) \prec_w (f(y_1), f(y_2), \cdots, f(y_n))$ for any increasing and convex function $f : I\!R \to I\!R$;

(b) if $\boldsymbol{x} \prec_w \boldsymbol{y}$ and $\boldsymbol{u} \prec_w \boldsymbol{v}$, then the concatenation of the vectors preserves the weak majorization: $(\boldsymbol{x}, \boldsymbol{u}) \prec_w (\boldsymbol{y}, \boldsymbol{v})$;

(c) assume that $x_1 = x_2 = \cdots = x_n = x$. Then $\boldsymbol{x} \prec_w \boldsymbol{y}$ if and only if $nx \leq \sum_{i=1}^{n} y_i$;

(d) assume that $m = n$ and $x_1 = x_2 = \cdots = x_n$, $u_1 = u_2 = \cdots = u_n$. If $\boldsymbol{x} \prec_w \boldsymbol{y}$ and $\boldsymbol{u} \prec_w \boldsymbol{v}$, then $\boldsymbol{x} + \boldsymbol{u} \prec_w \boldsymbol{y} + \boldsymbol{v}$.

A stochastic order that is directly related to the weak majorization is defined in Marshall and Olkin [12] between random vectors. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random variables, we have

$$\boldsymbol{X} \leq_{\mathrm{E}_3^{\uparrow}} \boldsymbol{Y} \text{ if } E[\phi(\boldsymbol{X})] \leq E[\phi(\boldsymbol{Y})],$$

for all ϕ of the form $\phi(\boldsymbol{X}) = \sum_{k=1}^{n} g(x_k)$ where $g : I\!R \to I\!R$ is increasing and convex.

Another stochastic order that will be used in the paper is the increasing convex order (cf. Stoyan [13]): Two random variables $X, Y \in I\!R$ are ordered by the increasing convex order, say $X \leq_{\mathrm{icx}} Y$, if $Ef(X) \leq Ef(Y)$ for any increasing and convex function $f : I\!R \to I\!R$, provided the expectations exist.

3 Smoothing Effect of the Superposition of Fluid Sources

In this section we shall only consider fluid traffic model. The extension of the results to models of discrete arrivals will be addressed in the next section.

3.1 Queueing Model and Evolution Equations

Consider a fluid queueing network model consisting of K stations, each equipped with a server and an infinite-capacity buffer. The service rate of server k at time t is $c_k(t)$, $1 \le k \le K$, $t \ge 0$. These rates can vary in time in an arbitrary way. Fluids arrive at station 1 and traverse the stations in the order of $1, 2, \ldots, K$. The (aggregated) arrival rate at time t is $a(t)$, $t \ge 0$.

There are (constant) propagations delays, denoted by $\delta_2, \delta_3, \cdots, \delta_K$, between the successive stations. Traffic leaving station $k-1$ arrives δ_k time units later at station k, $2 \le k \le K$. Let $\delta_1 := 0$, and for $1 \le k \le K$, $\Delta_k := \sum_{i=1}^{k} \delta_i$ denotes the total propagation delay from station 1 to station k.

The tandem queueing network is illustrated in Figure 1.

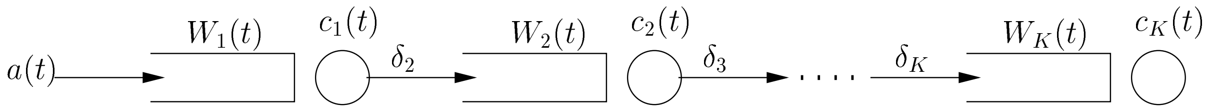

Figure 1: Tandem Queueing Network with Propagation Delays

Note that the variable service rates allow one to model situations where there are background traffic and cross traffic through the stations.

For simplicity of presentation, we shall use a slotted model and assume that arrival and service rates change only at integer time epochs, and that these rate processes are right-continuous. For any integer t, let $A(t)$ be the total amount of fluid that arrives during time interval $[t, t+1)$, and $C_k(t)$ the total potential service during time interval $[t, t+1)$ at station k, i.e., $C_k(t) = \int_t^{t+1} c_k(s)ds$. We further assume that the propagation delays are all integers.

Let $W_k(t)$ be the unfinished work in queue k at time t, $1 \le k \le K$, $t \ge 0$. The initial workload (seen by fluid arriving at time 0) at server k is denoted by $W_k(\Delta_k)$, $1 \le k \le K$. We shall be interested in the delayed workload $W_k(t + \Delta_k)$ and will denote $\widetilde{W}_k(t) \equiv W_k(t + \Delta_k)$. Similarly, we denote $\widetilde{C}_k(t) \equiv C_k(t + \Delta_k)$.

The following basic theorem establishes a recursion between the partial sums of the (delayed) unfinished works $\sum_{i=1}^{k} \widetilde{W}_i(t)$. The theorem is valid for a slightly more general queueing model where the arrival rate is decreasing (instead of constant) within time intervals $[t, t+1)$ for any integer t.

Theorem 3 *Assume that the arrival rate $a(s)$ is decreasing (in the sense of nonincreasing) and the service rates $c_k(s)$ are constant within time intervals $[t, t+1)$ for integer t. Then the delayed workloads $\widetilde{W}_k(t)$ satisfy the following recursive equation for integers $t = 0, 1, 2, \ldots$ and for $k = 1, 2, \ldots, K$:*

$$\sum_{i=1}^{k} \widetilde{W}_i(t+1)$$
$$= \max\left(\sum_{i=1}^{k} \widetilde{W}_i(t) + \sum_{i=1}^{k-1} \widetilde{C}_i(t) + A(t),\ \sum_{i=1}^{k-1} \widetilde{W}_i(t+1) + \sum_{i=1}^{k} \widetilde{C}_i(t)\right) - \sum_{i=1}^{k} \widetilde{C}_i(t). \tag{2}$$

Proof. Let us first consider the case $k = 1$. It is easy to see that the following differential equation holds for any $t \in I\!R$:

$$\dot{W}_1(t) = a(t) - c_1(t), \qquad W_1(t) > 0.$$

Consider any interval $[t, t+1)$ for $t \in I\!N$. Since $a(s)$ is decreasing and $c_1(s)$ is constant for $s \in [t, t+1)$, if the unfinished work goes to zero within the time interval $[t, t+1)$, it remains zero until time $t+1$. Thus, for any $t \in I\!N$ and any $0 < \epsilon \leq 1$:

$$W_1(t+\epsilon) = \max(W_1(t) + \int_0^{\epsilon} a(t+s)ds - \int_0^{\epsilon} c_1(t+s)ds, 0).$$

Hence, for any $t \in I\!N$,

$$W_1(t+1) = \max(W_1(t) + A(t) - C_1(t), 0) = \max(W_1(t) + A(t), C_1(t)) - C_1(t), \tag{3}$$

so that (2) holds for $k = 1$.

Consider now equation (2) for the general case of k. For $k = 1, 2, \ldots, K$, let $a_k(t)$ be the arrival rate at station k and $d_k(t)$ the departure rate from station k. Let $A_k(t)$ be the total fluid arrival at station k and $D_k(t)$ be the total fluid departure from station k during the time interval $[t, t+1)$. By definition, $a_1(t) \equiv a(t)$ and $A_1(t) \equiv A(t)$, and for all $k = 2, \ldots, K$,

$$a_k(t+\delta_k) = d_{k-1}(t), \quad A_k(t+\delta_k) = D_{k-1}(t). \tag{4}$$

It is easy to see that for any $s \geq 0$,

$$d_1(s) = \begin{cases} c_1(s), & W_1(s) > 0; \\ a_1(s), & W_1(s) = 0. \end{cases} \tag{5}$$

Note that $W_1(s) = 0$ implies $a_1(s) \leq c_1(s)$. Thus, the departure rate process $d_1(s)$ is decreasing for $s \in [t, t+1)$, where $t \in I\!N$.

For any integer $t \in I\!N$, if $W_1(s) > 0$ for all $s \in [t, t+1)$, then $D_1(t) = C_1(t)$; in which case, $W_1(t) + A_1(t) > C_1(t)$. Otherwise, if $W_1(s) = 0$ for some $s \in [t, t+1)$, then $D_1(t) = W_1(t) + A_1(t)$; in which case, $W_1(t) + A_1(t) \leq C_1(t)$. Thus, for any integer $t \in I\!N$,

$$D_1(t) = \min\left(W_1(t) + A_1(t), C_1(t)\right). \tag{6}$$

Owing to the fact that the departure rate process $d_1(s)$ is decreasing for $s \in [t, t+1)$, one can use a simple recursion to extend the relations (3), (5), and (6) to an arbitrary station k. That is, for any $1 \leq k \leq K$ and any $t \in I\!N$ and $t \leq s < t+1$:

$$\begin{aligned} W_k(t+1) &= \max(W_k(t) + A_k(t), C_k(t)) - C_k(t), && (7) \\ d_k(s) &= \begin{cases} c_k(s), & W_k(s) > 0; \\ a_k(s), & W_k(s) = 0. \end{cases} && (8) \\ D_k(t) &= \min\left(W_k(t) + A_k(t), C_k(t)\right). && (9) \end{aligned}$$

We now show the following relation for any k and j such that $1 \leq k \leq K$ and $1 \leq j \leq k$:

$$\sum_{i=j}^{k} \widetilde{W}_i(t+1) = \max\left(\sum_{i=j}^{k} \widetilde{W}_i(t) + \sum_{i=j}^{k-1} \widetilde{C}_i(t) + \widetilde{A}_j(t),\ \sum_{i=j}^{k-1} \widetilde{W}_i(t+1) + \sum_{i=j}^{k} \widetilde{C}_i(t)\right) - \sum_{i=j}^{k} \widetilde{C}_i(t), \tag{10}$$

where $\widetilde{A}_j(t) \equiv A_j(t + \Delta_j)$. It is clear that relation (2) is a simple consequence of (10).

We arbitrarily fix $1 \leq k \leq K$, and show (10) by backward induction on j using relations (7), (9) and (4). The induction basis $j = k$ holds, thanks to (7). Assume (10) holds for some $2 \leq j \leq k$. We consider $j-1$. Using the inductive assumption and (7), we obtain

$$\sum_{i=j-1}^{k} \widetilde{W}_i(t+1)$$

$$
\begin{aligned}
&= \max\left(\sum_{i=j}^{k}\widetilde{W}_i(t)+\sum_{i=j}^{k-1}\widetilde{C}_i(t)+\widetilde{A}_j(t)+\widetilde{W}_{j-1}(t+1),\right.\\
&\qquad \left.\sum_{i=j}^{k-1}\widetilde{W}_i(t+1)+\widetilde{W}_{j-1}(t+1)+\sum_{i=j}^{k}\widetilde{C}_i(t)\right)-\sum_{i=j}^{k}\widetilde{C}_i(t)\\
&= \max\left(\sum_{i=j}^{k}\widetilde{W}_i(t)+\sum_{i=j}^{k-1}\widetilde{C}_i(t)+\widetilde{A}_j(t)+\max\left(\widetilde{W}_{j-1}(t)+\widetilde{A}_{j-1}(t),\widetilde{C}_{j-1}(t)\right)-\widetilde{C}_{j-1}(t),\right.\\
&\qquad \left.\sum_{i=j-1}^{k-1}\widetilde{W}_i(t+1)+\sum_{i=j}^{k}\widetilde{C}_i(t)\right)-\sum_{i=j}^{k}\widetilde{C}_i(t)\\
&= \max\left(\sum_{i=j}^{k}\widetilde{W}_i(t)+\sum_{i=j}^{k-1}\widetilde{C}_i(t)+\widetilde{A}_j(t)+\max\left(\widetilde{W}_{j-1}(t)+\widetilde{A}_{j-1}(t),\widetilde{C}_{j-1}(t)\right),\right.\\
&\qquad \left.\sum_{i=j-1}^{k-1}\widetilde{W}_i(t+1)+\sum_{i=j-1}^{k}\widetilde{C}_i(t)\right)-\sum_{i=j-1}^{k}\widetilde{C}_i(t) \qquad (11)
\end{aligned}
$$

Using (4) and (9), we have

$$
\begin{aligned}
\widetilde{A}_j(t) &= A_j(t+\Delta_j)=A_j(t+\Delta_{j-1}+\delta_j)=D_{j-1}(t+\Delta_{j-1})\\
&= \min\left(W_{j-1}(t+\Delta_{j-1})+A_{j-1}(t+\Delta_{j-1}),C_{j-1}(t+\Delta_{j-1})\right)\\
&= \min\left(\widetilde{W}_{j-1}(t)+\widetilde{A}_{j-1}(t),\widetilde{C}_{j-1}(t)\right) \qquad (12)
\end{aligned}
$$

Using in addition the fact that $\min(X,Y)+\max(X,Y)=X+Y$, we obtain from relations (11) and (12) that

$$
\begin{aligned}
&\sum_{i=j-1}^{k}\widetilde{W}_i(t+1)\\
&\quad= \max\left(\sum_{i=j}^{k}\widetilde{W}_i(t)+\sum_{i=j}^{k-1}\widetilde{C}_i(t)+\widetilde{W}_{j-1}(t)+\widetilde{A}_{j-1}(t)+\widetilde{C}_{j-1}(t),\right.\\
&\qquad\qquad \left.\sum_{i=j-1}^{k-1}\widetilde{W}_i(t+1)+\sum_{i=j-1}^{k}\widetilde{C}_i(t)\right)-\sum_{i=j-1}^{k}\widetilde{C}_i(t)\\
&\quad= \max\left(\sum_{i=j-1}^{k}\widetilde{W}_i(t)+\sum_{i=j-1}^{k-1}\widetilde{C}_i(t)+\widetilde{A}_{j-1}(t),\right.
\end{aligned}
$$

$$\left.\sum_{i=j-1}^{k-1} \widetilde{W}_i(t+1) + \sum_{i=j-1}^{k} \widetilde{C}_i(t)\right) - \sum_{i=j-1}^{k} \widetilde{C}_i(t) \tag{13}$$

Hence, by induction, relation (10) holds for any $1 \leq j \leq k$.

This completes the proof of the theorem. ∎

3.2 Monotonicity of the Traffic Smoothness in the Number of Sources

Let there be N statistically identical (homogeneous) and independent sources which are described by their rate processes $a^i(t)$, $1 \leq i \leq N$, $t \geq 0$. We consider the tandem queueing network described in the previous subsection which is fed with the *normalized* aggregate traffic whose rate process is defined by

$$a(t) = \frac{1}{N} \sum_{i=1}^{N} a^i(t) \tag{14}$$

We assume, as before, that the rate processes change state only at integer time epochs.

Consider now another tandem queueing network which differs from the previous one only in the input traffic. Instead of aggregating N sources, the input traffic is the superposition of $N-1$ statistically identical and independent sources with rate processes $a^i(t)$, $1 \leq i \leq N-1$, $t \geq 0$. The *normalized* aggregate traffic has the rate process

$$a'(t) = \frac{1}{N-1} \sum_{i=1}^{N-1} a^i(t) \tag{15}$$

Let $W_k^N(t)$ and $\widetilde{W}_k^N(t)$, $1 \leq k \leq K$, be the workload (and delayed workload) processes defined with the aggregation of N sources $a(t)$, and $W_k^{N-1}(t)$ and $\widetilde{W}_k^{N-1}(t)$, $1 \leq k \leq K$, the workload (and delayed workload) processes defined with the aggregation of $N-1$ sources $a'(t)$,

We shall compare the two systems in terms of the (partial) sum of delayed workloads. Note that due to the normalization (by N in the case of N sources and by $N-1$ in the case of $N-1$ sources), the two aggregate sources have the same intensity. Our main result is stated in the following theorem.

Theorem 4 *Assume that the arrival and service rate processes $a^n(t)$, $1 \le n \le N$, $c_k(t)$, $1 \le k \le K$, are mutually independent, but are otherwise arbitrary. The aggregate traffic of N homogeneous sources is smoother than that of $N-1$ sources in the sense that for any $t \in I\!N$ and any k, $1 \le k \le K$,*

$$\sum_{i=1}^{k} \widetilde{W}_i^N(t) \le_{\rm icx} \sum_{i=1}^{k} \widetilde{W}_i^{N-1}(t), \tag{16}$$

provided that initial workloads in the two systems are identical: $W_k^N(0) = W_k^{N-1}(0)$ for all $1 \le k \le K$.

Proof. We refer to the tandem queueing network fed by N sources as the W^N-system, and that fed by $N-1$ sources as the W^{N-1}-system.

We consider two other queueing systems, each consisting of N parallel tandem queueing networks which are each identical to the above tandem queueing network, with the same propagation delays $\delta_2, \cdots, \delta_K$, the same service rate processes $c_1(t), \cdots, c_K(t)$, and the same initial workloads $W_1^N(0), \cdots, W_K^N(0)$. We label the tandem queueing networks as $1, 2, \ldots, N$.

In the first system, referred to as R-system, at any time t, each source, say source i, simultaneously sends traffic at rate $a^i(t)/N$ to the N parallel tandem queueing networks. Figure 2(a) illustrates the input traffic scheme of these tandem queues.

In the S-system, each source, say source i, simultaneously sends traffic at rate $a^i(t)/(N-1)$ to the $N-1$ tandem queueing networks $1, \ldots, i-1, i+1, \ldots, N$. Thus each tandem queueing network is fed by $N-1$ sources. Figure 2(b) illustrates the input traffic scheme of S-system.

Let $R_k^n(t)$ and $\widetilde{R}_k^n(t)$ (respectively $S_k^n(t)$ and $\widetilde{S}_k^n(t)$), $1 \le n \le N$, $1 \le k \le K$, be the workload and the delayed workload processes of the k-th station of the n-th tandem queueing network of the R-system (respectively S-system).

It is clear that the N parallel tandem queueing networks in the R-system are identical and that they are stochastically identical to the W^N-system:

$$R_k^n(t) = W_k^N(t), \quad \widetilde{R}_k^n(t) = \widetilde{W}_k^N(t), \qquad t \ge 0,\ 1 \le n \le N,\ 1 \le k \le K. \tag{17}$$

Since the N sources are stochastically identical and arrivals to each tandem queueing system in the S-system consist of equal shares from each of $N-1$ sources, all N parallel

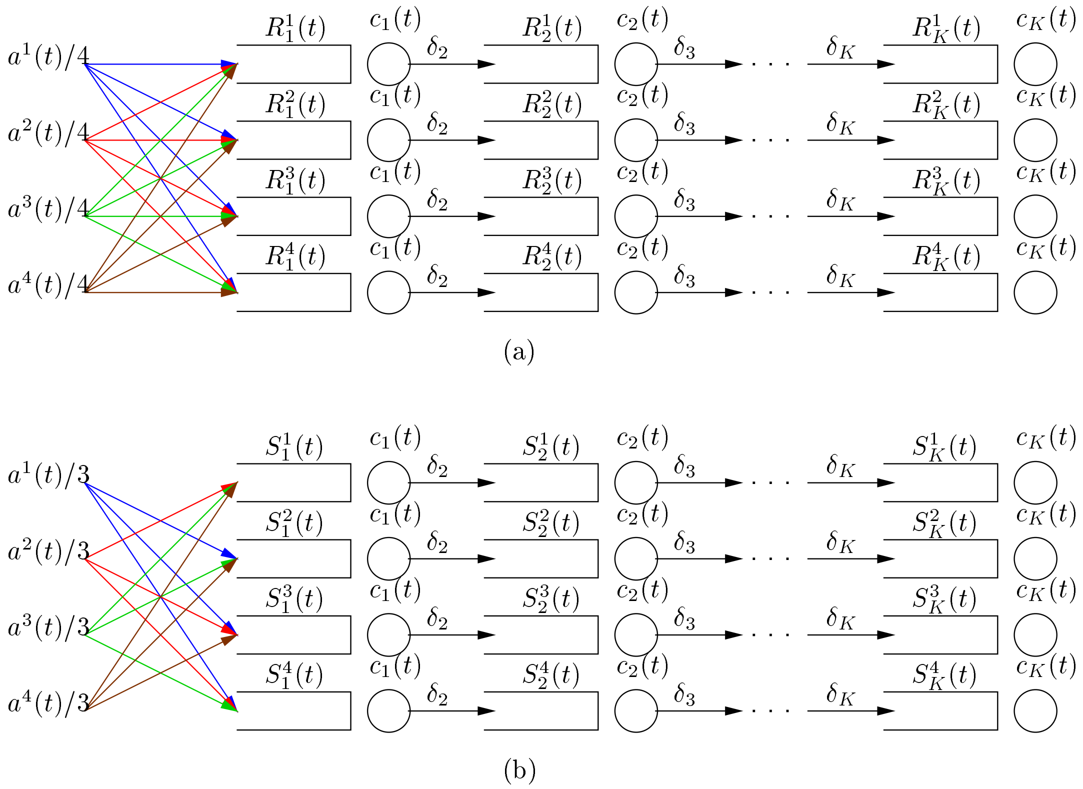

Figure 2: Parallel Tandem Queueing Networks. (a) The R-system corresponding to N sources. (b) The S-system corresponding to $N-1$ sources.

tandem queueing networks are stochastically identical and are stochastically the same as the W^{N-1}-system:

$$\left(S_1^n(t), \cdots, S_K^n(t)\right) =_d \left(W_1^{N-1}(t), \cdots, W_K^{N-1}(t)\right), \qquad 1 \le n \le N,\ t \ge 0, \tag{18}$$

$$\left(\widetilde{S}_1^n(t), \cdots, \widetilde{S}_K^n(t)\right) =_d \left(\widetilde{W}_1^{N-1}(t), \cdots, \widetilde{W}_K^{N-1}(t)\right), \qquad 1 \le n \le N,\ t \ge 0, \tag{19}$$

where $=_d$ denotes equality in distribution.

We shall show in the sequel that for any $t \ge 0$ and k, $1 \le k \le K$, the following stochastic order holds

$$\left(\sum_{i=1}^k \widetilde{R}_i^1(t), \cdots, \sum_{i=1}^k \widetilde{R}_i^N(t)\right) \le_{\mathrm{E}_3^\uparrow} \left(\sum_{i=1}^k \widetilde{S}_i^1(t), \cdots, \sum_{i=1}^k \widetilde{S}_i^N(t)\right) \tag{20}$$

Once this is proven, identities (17) and (19) will allow us to conclude (16). Indeed, for all increasing and convex function $f : I\!R \to I\!R$, we obtain from (20) that

$$\begin{aligned} N \times E\left[f\left(\sum_{i=1}^k \widetilde{W}_i^N(t)\right)\right] &= \sum_{n=1}^N E\left[f\left(\sum_{i=1}^k \widetilde{R}_i^n(t)\right)\right] \\ &\le \sum_{n=1}^N E\left[f\left(\sum_{i=1}^k \widetilde{S}_i^n(t)\right)\right] = N \times E\left[f\left(\sum_{i=1}^k \widetilde{W}_i^{N-1}(t)\right)\right] \end{aligned}$$

so that

$$E\left[f\left(\sum_{i=1}^k \widetilde{W}_i^N(t)\right)\right] \le E\left[f\left(\sum_{i=1}^k \widetilde{W}_i^{N-1}(t)\right)\right].$$

In order to prove (20), we arbitrarily fix the N arrival rate processes $a^n(t)$, $1 \le n \le N$, and couple them in the two systems. Similarly, we arbitrarily fix the K service rate processes $c_k(t)$, $1 \le k \le K$, and coupled them in all of the $2N$ tandem queueing networks.

Under such a coupling, we are going to show that for all t, we have the following weak majorization ordering:

$$\left(\sum_{i=1}^k \widetilde{R}_i^1(t), \cdots, \sum_{i=1}^k \widetilde{R}_i^N(t)\right) \prec_w \left(\sum_{i=1}^k \widetilde{S}_i^1(t), \cdots, \sum_{i=1}^k \widetilde{S}_i^N(t)\right) \tag{21}$$

The proof of this last relation relies on Theorem 3 and Corollary 2.

Applying Theorem 3 to the R-system and S-system implies that the following recursions hold for any n, $1 \leq n \leq N$:

$$\begin{aligned}
&\sum_{i=1}^{k} \widetilde{R}_i^n(t+1) \\
&= \max\left(\sum_{i=1}^{k} \widetilde{R}_i^n(t) + \sum_{i=1}^{k-1} \widetilde{C}_i(t) + A(t),\ \sum_{i=1}^{k-1} \widetilde{R}_i^n(t+1) + \sum_{i=1}^{k} \widetilde{C}_i(t)\right) - \sum_{i=1}^{k} \widetilde{C}_i(t), \qquad (22)\\
&\sum_{i=1}^{k} \widetilde{S}_i^n(t+1) \\
&= \max\left(\sum_{i=1}^{k} \widetilde{S}_i^n(t) + \sum_{i=1}^{k-1} \widetilde{C}_i(t) + B^n(t),\ \sum_{i=1}^{k-1} \widetilde{S}_i^n(t+1) + \sum_{i=1}^{k} \widetilde{C}_i(t)\right) - \sum_{i=1}^{k} \widetilde{C}_i(t), \qquad (23)
\end{aligned}$$

where $A(t)$ is the cumulative arrivals in $[t, t+1)$ corresponding to the rate process $a(t)$ defined in (14), $B^n(t)$ is the cumulative arrivals in $[t, t+1)$ of the aggregate traffic of the n-th tandem queueing network in S-system.

Let $b^n(t)$ be the rate process of the aggregate traffic of the n-th tandem queueing network in S-system. It is easy to see that

$$\sum_{n=1}^{N} b^n(t) = \sum_{n=1}^{N} (N-1)\frac{a^n(t)}{N-1} = \sum_{n=1}^{N} a^n(t) = Na(t),$$

so that

$$\sum_{n=1}^{N} B^n(t) = NA(t).$$

Thus, owing to Corollary 2, for any t,

$$(A(t), \cdots, A(t)) \prec_w (B^1(t), \cdots, B^N(t)). \qquad (24)$$

We now use a simple induction on (k, t) to show (21). The induction basis with $t = 0$ is simple. Since all the tandem queueing networks have the same initial workloads $W_1^N(0), \cdots, W_K^N(0)$, we have for $k = 1, 2, \cdots, K$:

$$\left(\sum_{i=1}^{k} \widetilde{R}_i^1(0), \cdots, \sum_{i=1}^{k} \widetilde{R}_i^N(0)\right)$$

$$= \left(\sum_{i=1}^{k} \widetilde{S}_i^1(0), \cdots, \sum_{i=1}^{k} \widetilde{S}_i^N(0) \right)$$

Assume (21) holds for some $t \geq 0$. We show that it holds for $t+1$ by induction on k. For $k = 1$, we obtain from (22) and (23) that

$$\begin{aligned} \widetilde{R}_1^n(t+1) &= \max\left(\widetilde{R}_1^n(t) + A(t), \widetilde{C}_1(t)\right) - \widetilde{C}_1(t), \\ \widetilde{S}_1^n(t+1) &= \max\left(\widetilde{S}_1^n(t) + B^n(t), \widetilde{C}_1(t)\right) - \widetilde{C}_1(t). \end{aligned}$$

Using the inductive assumption we have

$$\left(\widetilde{R}_1^1(t), \cdots, \widetilde{R}_1^N(t)\right) \prec_w \left(\widetilde{S}_1^1(t), \cdots, \widetilde{S}_1^N(t)\right),$$

which, together with (24), allow us to apply Corollary 2 yielding

$$\left(\widetilde{R}_1^1(t) + A(t), \cdots, \widetilde{R}_1^N(t) + A(t)\right) \prec_w \left(\widetilde{S}_1^1(t) + B^1(t), \cdots, \widetilde{S}_1^N(t) + B^N(t)\right).$$

Recall that the $\max(\cdot)$ operator is an increasing and convex function. Thus, using again Corollary 2, we obtain from the above majorization relations that

$$\left(\widetilde{R}_1^1(t+1), \cdots, \widetilde{R}_1^N(t+1)\right) \prec_w \left(\widetilde{S}_1^1(t+1), \cdots, \widetilde{S}_1^N(t+1)\right).$$

Assume (21) holds for $t+1$ and some $k \geq 1$. Then,

$$\begin{aligned} \left(\sum_{i=1}^{k+1} \widetilde{R}_i^1(t), \cdots, \sum_{i=1}^{k+1} \widetilde{R}_i^N(t) \right) &\prec_w \left(\sum_{i=1}^{k+1} \widetilde{S}_i^1(t), \cdots, \sum_{i=1}^{k+1} \widetilde{S}_i^N(t) \right) \\ \left(\sum_{i=1}^{k} \widetilde{R}_i^1(t+1), \cdots, \sum_{i=1}^{k} \widetilde{R}_i^N(t+1) \right) &\prec_w \left(\sum_{i=1}^{k} \widetilde{S}_i^1(t+1), \cdots, \sum_{i=1}^{k} \widetilde{S}_i^N(t+1) \right), \end{aligned}$$

which, in addition to (24), yield, cf. Corollary 2,

$$\begin{aligned} &\left(\sum_{i=1}^{k+1} \widetilde{R}_i^1(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + A(t), \cdots, \sum_{i=1}^{k+1} \widetilde{R}_i^N(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + A(t) \right) \\ \prec_w &\left(\sum_{i=1}^{k+1} \widetilde{S}_i^1(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + B^1(t), \cdots, \sum_{i=1}^{k+1} \widetilde{S}_i^N(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + B^N(t) \right) \end{aligned}$$

$$\left(\sum_{i=1}^{k} \widetilde{R}_i^1(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t), \cdots, \sum_{i=1}^{k} \widetilde{R}_i^N(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t)\right)$$
$$\prec_w \quad \left(\sum_{i=1}^{k} \widetilde{S}_i^1(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t), \cdots, \sum_{i=1}^{k} \widetilde{S}_i^N(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t)\right).$$

Since we have from (22) and (23) that

$$\sum_{i=1}^{k+1} \widetilde{R}_i^n(t+1)$$
$$= \max\left(\sum_{i=1}^{k+1} \widetilde{R}_i^n(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + A(t),\ \sum_{i=1}^{k} \widetilde{R}_i^n(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t)\right) - \sum_{i=1}^{k+1} \widetilde{C}_i(t),$$
$$\sum_{i=1}^{k+1} \widetilde{S}_i^n(t+1)$$
$$= \max\left(\sum_{i=1}^{k+1} \widetilde{S}_i^n(t) + \sum_{i=1}^{k} \widetilde{C}_i(t) + B^n(t),\ \sum_{i=1}^{k} \widetilde{S}_i^n(t+1) + \sum_{i=1}^{k+1} \widetilde{C}_i(t)\right) - \sum_{i=1}^{k+1} \widetilde{C}_i(t),$$

an application of Corollary 2 allows us to conclude that

$$\left(\sum_{i=1}^{k+1} \widetilde{R}_i^1(t+1), \cdots, \sum_{i=1}^{k+1} \widetilde{R}_i^N(t+1)\right) \prec_w \left(\sum_{i=1}^{k+1} \widetilde{S}_i^1(t+1), \cdots, \sum_{i=1}^{k+1} \widetilde{S}_i^N(t+1)\right).$$

Therefore, by induction, (21) holds for all $t \geq 0$ and all $1 \leq k \leq K$.

We then readily obtain the ordering (20) using Proposition 1 by unconditioning the source rate processes $a^n(t)$, $1 \leq n \leq N$, and the service rate processes $c_k(t)$, $1 \leq k \leq K$, as well as the choices of the queues to feed in S-system. This completes the proof. ∎

It is worthwhile noticing that the comparisons (20) and (21) between the R-system and S-system hold for any arbitrary sources $a^1(t), \cdots, a^N(t)$, even if they are not homogeneous.

4 Smoothing Effect of the Superposition of Discrete Sources

We now extend the results of the previous section to a discrete traffic model.

Consider a slotted queueing network model with discrete arrivals. The queueing network consists of K stations in tandem, numbered from 1 to K, each equipped with a server and an infinite-capacity buffer. At the beginning of the t-th time slot, $A(t)$ customers arrive at station 1. At the end of time slot t, at most $C_k(t)$ customers are served at station k, $1 \le k \le K$, $t \ge 0$. Customers leaving station $k-1$ at the end of slot t become available at the beginning of slot $t+1+\delta_k$ at station k, $2 \le k \le K$, where δ_k is the propagation delay. As in the previous section, we let $\delta_1 := 0$ by convention, and for $1 \le k \le K$, $\Delta'_k := (k-1) + \sum_{i=1}^{k} \delta_i$. Note that Δ'_k is larger than Δ_k used in the previous section because of the fact that service completions occur at the end of a time slot.

Let $Q_k(t)$ be the number of customers in queue k at the beginning of slot t, $1 \le k \le K$, $t \ge 0$. As before, we are interested in the delayed queue length $Q_k(t+\Delta'_k)$ and will denote $\widetilde{Q}_k(t) \equiv Q_k(t+\Delta'_k)$. Similarly, we denote $\widetilde{C}_k(t) \equiv C_k(t+\Delta'_k)$.

By mimicking the proof of Theorem 3 we can establish

Theorem 5 *Then variables $\widetilde{Q}_k(t)$ satisfy the following recursive equation for integers $t = 0, 1, 2, \ldots$ and for $k = 1, 2, \ldots, K$:*

$$\begin{aligned}
&\sum_{i=1}^{k} \widetilde{Q}_i(t+1) \\
&\quad = \max\left(\sum_{i=1}^{k} \widetilde{Q}_i(t) + \sum_{i=1}^{k-1} \widetilde{Q}_i(t) + A(t),\ \sum_{i=1}^{k-1} \widetilde{Q}_i(t+1) + \sum_{i=1}^{k} \widetilde{C}_i(t) \right) - \sum_{i=1}^{k} \widetilde{C}_i(t). \qquad (25)
\end{aligned}$$

Consider now the comparison of the smoothness of the aggregation of N sources against that of M sources, with $N > M$. The arrival processes of the sources are described by $A^n(t)$, $n = 1, 2, \cdots, N$. As before, we need to *normalize* the traffic so that the aggregation of N sources and that of M sources have the same traffic intensity. In order to do so, we assume that in the aggregation of N (resp. M) sources, each arrival is a batch of M (resp. N) customers. Thus, the (*normalized*) aggregate traffic of N sources is defined by

$$A(t) = \frac{1}{N} \sum_{n=1}^{N} M A^n(t) \qquad (26)$$

and that of M sources is defined by

$$B(t) = \frac{1}{M} \sum_{n=1}^{M} N A^n(t). \qquad (27)$$

Let $Q_k^N(t)$ and $\widetilde{Q}_k^N(t)$, $1 \le k \le K$, be the queue length and the delayed queue length sequences defined with the aggregation of N sources $A(t)$, and $Q_k^M(t)$ and $\widetilde{Q}_k^M(t)$, $1 \le k \le K$, the queue length and the delayed queue length sequences defined with the aggregation of M sources $B(t)$, We compare the two systems in terms of the (partial) sum of delayed queue lengths.

Theorem 6 *Assume that the arrival and service processes $A^n(t)$, $1 \le n \le N$, $C_k(t)$, $1 \le k \le K$, are mutually independent, but are otherwise arbitrary. The aggregate traffic of N homogeneous sources is smoother than that of M sources in the sense that for any $t \in I\!N$ and any k, $1 \le k \le K$,*

$$\sum_{i=1}^{k} \widetilde{Q}_i^N(t) \le_{\rm icx} \sum_{i=1}^{k} \widetilde{Q}_i^M(t), \tag{28}$$

provided that initial queue lengths of the two systems are identical: $Q_k^N(0) = Q_k^M(0)$ for all $1 \le k \le K$.

The proof of the above theorem is analogous to that of Theorem 4. We can still construct two queueing systems, each consists of N parallel tandem queueing networks which are all identical to the above tandem queueing network,

In the S-system, each source sends customers to M tandem queueing networks simultaneously, and each tandem queueing network is fed by M sources. The choice of feeding the queues is made randomly according to the uniform distribution and can be done in the following way. Source 1 chooses with equal probability M (over N) tandem queueing networks to send traffic to. A token is assigned to each of these M tandem queueing networks. The other sources subsequently choose with equal probability M tandem queueing networks among those with fewer than M tokens to send traffic to, and then tokens to these tandem queues.

The remaining arguments in the proof of Theorem 4 go through straightforward.

5 Concluding Remarks

In this paper we have analyzed the smoothing effect of superposing homogeneous sources in a network. We have considered a tandem queueing network representing the nodes that these sources pass through. The servers in the tandem queues have different and variable service rates. We have shown that for arbitrary arrival and service processes which are mutually independent, the sum of delayed unfinished works (in the fluid model) and the sum of delayed queue lengths (in the discrete model) in the tandem queues is monotone in the number of homogeneous sources in the increasing convex order sense, provided the total intensity of the aggregate traffic is constant.

In order to prove these results, we have established recursive equations governing the partial sums of the delayed unfinished works and of the delayed queue lengths. We believe that such equations would be useful for other qualitative and quantitative analyses of such systems. In particular, it is simple to show that the partial sums of the above mentioned state variables are monotone in the input traffic in stochastic ordering sense and in the increasing convex ordering sense, see [2] examples of systems exhibiting such properties.

References

[1] F. Baccelli, P. Bremaud, *Elements of Queueing Theory*, Springer-Verlag, Berlin, 1994.

[2] F. Baccelli, Z. Liu, "Comparison Properties of Stochastic Decision Free Petri Nets", *IEEE Trans. on Automatic Control*, Vol. 37, pp. 1905-1920, 1992.

[3] N. Bäuerle, "The advantage of small machines in a stochastic fluid production process", *Mathematical Methods of Operations Research*, 47:83–97, 1998.

[4] N. Bäuerle, "How to improve the performance of ATM multiplexers, *Operations Research Letters*, 24:81–89, 1999.

[5] D. Botvich, N. Duffield, "Large Deviations, the Shape of the Loss Curve, and Economics of Scale in Large Multiplexers", *Queueing Systems*, Vol. 20, pp. 293-320, 1995.

[6] C. Courcoubetis, R. Weber, "Buffer Overflow Asymptoticsfor a Buffer Handling Many Traffic Sources", *Journal of Applied Probability*, Vol. 33, pp. 886-903, 1996.

[7] V. Dumas, A. Simonian, "Asymptotic Bounds for the Fluid Queue Fed by Subexponential On-Off Sources", preprint.

[8] A. Elwalid, D. Mitra, R. Wentworth, "A New Approach for Allocating Buffers and Bandwidth to Heterogeneous Regulated Traffic in an ATM Node", *IEEE Journal on Selected Areas in Communications*, Vol 13, pp. 1105-1127, 1995.

[9] G. Koole, Z. Liu, "Stochastic Bounds for Queueing Systems with Multiple On-Off Sources", *Probability in the Engineering and Information Sciences*, Vol. 12, pp. 25-48, Jan. 1998.

[10] G. Koole, Z. Liu, D. Towsley, "Comparing Queueing Systems with Heterogeneous On-Off Sources", Technical Report WS-530, Vrije Universiteit Amsterdam, 1999.

[11] M. Mandjes, "A Note on Large Deviations for Small Buffers", preprint.

[12] A. W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

[13] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*. English translation (D.J. Daley editor), J. Wiley and Sons, New York, 1983.

[14] A. Weiss, "A New Technique for Analyzing Large Traffic Systems", *Advances in Applied Probability*, Vol. 18, pp. 506-532, 1986.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399