INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Limit laws for large product-form networks : connections with the Central Limit Theorem

Guy FAYOLLE - Jean-Marc LASGOUTTES

N° 2513
Mars 1995

PROGRAMME 1

Rapport de recherche

Limit laws for large product-form networks: connections with the Central Limit Theorem

Guy Fayolle * Jean-Marc Lasgouttes *

March 1995

Abstract

We consider a closed product-form network with n queues and m clients. We are interested in its asymptotic behaviour when m and n become simultaneously large. Our method relies on Berry-Esseen type approximations of the Central Limit Theorem. This leads to simple and natural conditions applicable to general networks, whereas the purely analytical methods used previously imposed restrictions on the queues. In particular, we show that the "optimal" dependence of m w.r.t. n is not necessarily linear. An application of these results to a transportation network is presented. We show how some queues can act as bottlenecks, limiting thus the efficiency of the whole system.

*Postal address: INRIA — Domaine de Voluceau, Rocquencourt, BP105 — 78153 Le Chesnay, France.

Lois limites pour des grands réseaux à forme produit : connexions avec le Théorème Central Limite

Guy Fayolle*   Jean-Marc Lasgouttes *

Mars 1995

Résumé

On s'intéresse au comportement d'un réseau fermé à forme produit comportant n files et m clients quand m et n tendent vers l'infini simultanément. Notre méthode s'apparente aux approximations liées au Théorème Central Limite. Elle conduit à des conditions simples et naturelles, applicables à des réseaux généraux, alors que les méthodes purement analytiques utilisées jusque ici imposaient des restrictions sur les files. En particulier, on montre que la dépendance « optimale » de m par rapport à n n'est pas nécessairement linéaire.

*INRIA — Domaine de Voluceau, Rocquencourt, BP105 — 78153 Le Chesnay, France.

1 Introduction

In many applications (telecommunications, transportation, etc.), it is desirable to understand the behaviour and performance of stochastic networks as their size increases. From an engineering point of view, the problem can be roughly formulated as follows.

> *Consider a closed network with* n *nodes and exactly* m *customers circulating inside. Find all functions* f*, such that* $m = f(n)$ *yields an* interesting *performance of the system as* n *increases.*

In this paper, we start from the famous so-called product-form networks, which play an important role in quantitative analysis of systems. Although the equilibrium state probabilities have then a simple expression (see for example Kelly [3]), non-trivial problems remain, due to an intrinsic combinatorial explosion in the formulas, especially in those involving the famous normalizing constant. To circumvent these drawbacks, the idea is to compute asymptotic expansions of the characteristic values of the network, when m and n both tend to infinity.

This approach has been considered by Knessl and Tier [4], Kogan and Birman [5, 6, 1] and Malyshev and Yakovlev [8]. However, they rely on purely analytical methods which are difficult to use in a general setting and do not really give a structural explanation of the phenomena involved.

The method proposed in this paper has direct connections with the *central limit theorem*: for instance, instead of considering the normalizing constant as an integral in the complex plane, we express it as the distribution of a scaled sum of independent random variables. Besides giving a clear interpretation of the computations, this allows to handle directly the general case of single-chain closed networks. More interestingly, the results can be interpreted as an insensitivity result: as the number of stations n and the number of customers $m(n)$ go to infinity, the network is equivalent to n independent queues, having the same characteristics and mean total number of customers $m(n)$. It turns out that the natural scaling parameter is indeed the flow λ_n in this equivalent network. As a consequence, we can state that $m(n)$ has *a priori* no theoretical reason of being proportional to n.

It is worth to emphasize that the problem of computing quantities (such as the above mentioned normalizing constant) appears frequently in statistical physics, in particular for Gibbs distributions in Markov random fields (see for example Whittle [14]). For the sake of completeness, it is of interest to quote the studies by Mitra, McKenna and Ramakrishnan [11, 9, 10, 12], where only the number of customers goes to infinity.

The present paper is organized as follows. The model is introduced in Section 2 and the main values of interest are expressed in terms of integrals of characteristic functions. In Section 3, asymptotics of these integrals are found in the case where the queues are in *normal* or *moderated traffic*, while Section 4 deals with the case where some queues are overloaded. These results are summarized in Section 5, and the choice of scalings is discussed. Section 6 is devoted to an application of these results to a service vehicle network (like the PRAXITÈLE project, now developed at INRIA); non quite intuitive results are enlightened. Section 7 contains some conclusive remarks and some bounds used in Sections 3 and 4 are proven in Appendix A.

2 Description of the model

We consider a closed BCMP network $\mathcal{B}_n$ with n queues and $m(n)$ clients. The number of clients at queue k at steady state is a random variable $Q_{k,n}$. The service rate at queue k when there are q_k customers is $\mu_{k,n}(q_k)$. The routing probability from queue k to queue ℓ is $p_{k,\ell,n}$ and P_n denotes the transition matrix supposed to be ergodic, with invariant measure $\pi_n = (\pi_{1,n}, \dots, \pi_{n,n})$, defined by:

$$\pi_n P_n = \pi_n \text{ and } \pi_{1,n} + \cdots + \pi_{n,n} = 1. \tag{2.1}$$

Then it is known that, for any $q_1, \dots, q_n \geq 0$ such that $q_1 + \cdots + q_n = m(n)$,

$$P_n(Q_{1,n} = q_1, \dots, Q_{n,n} = q_n) = Z^{-1}_{m(n),n} \prod_{k=1}^{n} \frac{\pi_{k,n}^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)}, \tag{2.2}$$

with the normalizing condition

$$Z_{m,n} = \sum_{q_1+\cdots+q_n=m} \prod_{k=1}^{n} \frac{\pi_{k,n}^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)}.$$

It is worth noting that our analysis applies to any network which has a product form equilibrium distribution like (2.2). For example, it includes, with a proper choice of the $\mu_{k,n}$'s, all systems having transition rates of the form $p_{k,\ell,n}\alpha_{k,n}(q_k)\beta_{\ell,n}(q_\ell)$. See Serfozo [13] for further examples.

For the sake of brevity, we shall write Z_n instead of $Z_{m(n),n}$, whenever non ambiguous. To each queue k, $1 \le k \le n$, we associate the generating function

$$f_{k,n}(z) \stackrel{\text{def}}{=} \sum_{q=1}^{\infty} \frac{z^q}{\mu_{k,n}(1)\cdots\mu_{k,n}(q)},$$

and we define

$$F_n(z) \stackrel{\text{def}}{=} \frac{1}{z^{m(n)}} \prod_{k=1}^{n} f_{k,n}(\pi_{k,n} z).$$

To the original closed network $\mathcal{B}_n$, we let correspond a new system $\mathcal{A}_n$ which is open and consists of n parallel queues, with service rates $\mu_{k,n}(x)$ and arrival intensity $\lambda_n \pi_{k,n}$ at queue k, where the choice of λ_n will be made more precise later. The queue length $X_{k,n}$ of the k's queue of $\mathcal{A}_n$ has a distribution given by

$$P(X_{k,n} = x) = \frac{(\lambda_n \pi_{k,n})^x}{f_{k,n}(\lambda_n \pi_{k,n}) \mu_{k,n}(1)\cdots\mu_{k,n}(x)},$$

and $X_{1,n}, \ldots, X_{n,n}$ are independent variables. We assume that $X_{k,n}$ has some finite moments of order $r \ge 2$ and introduce the following notation:

$$\begin{aligned}
m_{k,n} &\stackrel{\text{def}}{=} \mathbf{E}\, X_{k,n}, \\
\beta_{k,n}^{(r)} &\stackrel{\text{def}}{=} \mathbf{E}\, |X_{k,n} - m_{k,n}|^r, \\
\sigma_{k,n}^2 &\stackrel{\text{def}}{=} \operatorname{Var} X_{k,n} = \beta_{k,n}^{(2)}, \\
S_n &\stackrel{\text{def}}{=} X_{1,n} + \cdots + X_{n,n}, \\
\beta_n^{(r)} &\stackrel{\text{def}}{=} \beta_{1,n}^{(r)} + \cdots + \beta_{n,n}^{(r)}, \\
\sigma_n^2 &\stackrel{\text{def}}{=} \operatorname{Var} S_n = \beta_n^{(2)}.
\end{aligned}$$

Let $\varphi_{k,n}(\theta)$ be the characteristic function of $X_{k,n} - m_{k,n}$. Then, for any real θ,

$$\varphi_{k,n}(\theta) \stackrel{\text{def}}{=} \mathbf{E}\, e^{i(X_{k,n} - m_{k,n})\theta} = \frac{f_{k,n}(\pi_{k,n}\lambda_n e^{i\theta})}{f_{k,n}(\pi_{k,n}\lambda_n)} e^{-i m_{k,n}\theta}, \tag{2.3}$$

and

$$\varphi_n(\theta) \stackrel{\text{def}}{=} \mathbf{E}\, e^{i(S_n - \mathbf{E}\, S_n)\theta} = \varphi_{1,n}(\theta)\cdots\varphi_{n,n}(\theta) \tag{2.4}$$

The reason why $\mathcal{A}_n$ has been introduced is that the main performance characteristics of the network $\mathcal{B}_n$ can be expressed in terms of the distribution of $X_{1,n}, \ldots, X_{n,n}$, for a properly chosen value of λ_n:

Lemma 2.1 *For any choice of* $m(n)$, *there exists an unique* λ_n *such that*

$$\mathbf{E}\, S_n = \mathbf{E}[X_{1,n} + \cdots + X_{n,n}] = m(n). \tag{2.5}$$

With this choice of λ_n, we have

$$Z_n = \frac{F_n(\lambda_n)}{2\pi} \int_{-\pi}^{\pi} \prod_{k=1}^{n} \varphi_{k,n}(\theta) d\theta. \tag{2.6}$$

The mean number of clients in queue j *of* $\mathcal{B}_n$ *is given by*

$$\mathbf{E}\, Q_{j,n} = m_{j,n} - \frac{F_n(\lambda_n)}{2\pi Z_n} \int_{-\pi}^{\pi} i\varphi'_{j,n}(\theta) \prod_{k \neq j} \varphi_{k,n}(\theta) d\theta. \tag{2.7}$$

For any $\ell > 0$ *and* $q_1, \ldots, q_\ell \geq 0$, *the joint distribution of the length of queues* $1, \ldots, \ell$ *is*

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) = \tag{2.8}$$
$$\prod_{j=1}^{\ell} P(X_{j,n} = q_j) \frac{F_n(\lambda_n)}{2\pi Z_n} \int_{-\pi}^{\pi} e^{i \sum_{j=1}^{\ell}(q_j - m_{j,n})\theta} \prod_{k=\ell+1}^{n} \varphi_{k,n}(\theta) d\theta.$$

Proof One gets immediately from (2.3)

$$\begin{aligned}
m_{k,n} &= \pi_{k,n}\lambda_n \frac{f'_{k,n}(\pi_{k,n}\lambda_n)}{f_{k,n}(\pi_{k,n}\lambda_n)}, \\
\sigma^2_{k,n} &= \pi_{k,n}\lambda_n \frac{f'_{k,n}(\pi_{k,n}\lambda_n)}{f_{k,n}(\pi_{k,n}\lambda_n)} \\
&\quad + \pi^2_{k,n}\lambda^2_n \frac{f''_{k,n}(\pi_{k,n}\lambda_n) f_{k,n}(\pi_{k,n}\lambda_n) - f'^2_{k,n}(\pi_{k,n}\lambda_n)}{f^2_{k,n}(\pi_{k,n}\lambda_n)},
\end{aligned}$$

so that

$$\frac{\partial}{\partial \lambda_n} \mathbf{E}\, S_n = \frac{\sigma^2_{1,n} + \cdots + \sigma^2_{n,n}}{\lambda_n} > 0.$$

The mean number of clients in $\mathcal{A}_n$ is thus a strictly increasing function of λ_n, which equals zero when $\lambda_n = 0$ and goes to infinity with λ_n. This proves the first assertion of the lemma.

For any fixed n we consider the generating function

$$\begin{aligned}
\sum_{m=0}^{\infty} z^m Z_{m,n} &= \sum_{m=0}^{\infty} \sum_{q_1+\cdots+q_n=m} \prod_{k=1}^{n} \frac{(\pi_{k,n} z)^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)} \\
&= \sum_{q_1,\ldots,q_n} \prod_{k=1}^{n} \frac{(\pi_{k,n} z)^{q_k}}{\mu_{k,n}(1) \cdots \mu_{k,n}(q_k)} \\
&= \prod_{k=1}^{n} f_{k,n}(\pi_{k,n} z).
\end{aligned}$$

Since $m(n)$ is finite, λ_n is in the domain of convergence of this function, and Cauchy formula on the circle $\mathcal{C}$ with center 0 and radius λ_n yields

$$\begin{aligned}
Z_n &= \frac{1}{2i\pi}\int_{\mathcal{C}} \frac{\prod_{k=1}^{n} f_{k,n}(\pi_{k,n}z)}{z^{m(n)+1}}dz \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\prod_{k=1}^{n} f_{k,n}(\pi_{k,n}\lambda_n e^{i\theta})}{\lambda_n^{m(n)}} e^{-im(n)\theta}d\theta \qquad (2.9)\\
&= \frac{\prod_{k=1}^{n} f_{k,n}(\pi_{k,n}\lambda_n)}{\lambda_n^{m(n)}}\frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{k=1}^{n}\frac{f_{k,n}(\pi_{k,n}\lambda_n e^{i\theta})}{f_{k,n}(\pi_{k,n}\lambda_n)}e^{-im(n)\theta}d\theta,
\end{aligned}$$

which, taking (2.3) into account, amounts to Equation (2.6).

It is well known that

$$\mathbf{E}\,Q_{j,n} = \frac{1}{Z_n}\pi_{j,n}\frac{\partial Z_n}{\partial \pi_{j,n}}.$$

From (2.9), we get

$$\begin{aligned}
\mathbf{E}\,Q_{j,n} &= \frac{1}{2\pi Z_n}\int_{-\pi}^{\pi} \pi_{j,n}\lambda_n e^{i\theta} f'_{j,n}(\pi_{j,n}\lambda_n e^{i\theta})\frac{\prod_{k\neq j} f_{k,n}(\pi_{k,n}\lambda_n e^{i\theta})}{(\lambda_n e^{i\theta})^{m(n)}}d\theta,\\
&= \frac{F_n(\lambda_n)}{2\pi Z_n}\int_{-\pi}^{\pi}[m_{j,n}\varphi_{j,n}(\theta) - i\varphi'_{j,n}(\theta)]\prod_{k\neq j}\varphi_{k,n}(\theta)d\theta,
\end{aligned}$$

and (2.7) follows.

For any $\ell \geq 1$, the joint distributions take the form

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) = \frac{1}{Z_n}\prod_{j=1}^{\ell}\frac{\pi_{j,n}^{q_j}}{\mu_{j,n}(1)\cdots\mu_{j,n}(q_j)}Z^*,$$

where Z^* is the normalizing constant corresponding to the network of queues $\ell+1, \ldots, n$ with $m(n) - \sum_{j=1}^{\ell} q_j = m(n) - \sum q_j$ clients. Following the same lines as above, we obtain (2.8), since

$$\begin{aligned}
&P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell)\\
&\quad = \frac{1}{2\pi Z_n}\prod_{j=1}^{\ell}\frac{\pi_{j,n}^{q_j}}{\mu_{j,n}(1)\cdots\mu_{j,n}(q_j)}\int_{-\pi}^{\pi}\frac{\prod_{k=\ell+1}^{n} f_{k,n}(\pi_{k,n}\lambda_n e^{i\theta})}{(\lambda_n e^{i\theta})^{m(n)-\sum q_j+1}}d\theta.
\end{aligned}$$

■

These formulas are used in Sections 3 and 4 to derive asymptotic expansions. These expansions are given in terms of the operators O and Ω defined as follows:

$$a(\eta) = O(b(\eta)) \text{ iff } \exists K > 0,\ \forall \eta,\ |a(\eta)| \leq K|b(\eta)|,$$

$$a(\eta) = \Omega(b(\eta)) \text{ iff } a(\eta) = O(b(\eta)) \text{ and } b(\eta) = O(a(\eta)),$$

where η is unspecified argument. Unless otherwise stated, all these bounds are uniform with respect to n and all queue indexes.

3 Asymptotic expansion for normal traffic

In this section, we compute estimates of several performance measures of $\mathcal{B}_n$ when the queues are not saturated (in a sense that will be made precise in Lemma 3.1). The method that we use to compute our estimates is related to the Central Limit Theorem or, more precisely, to Berry-Esseen type expansions. The results of next lemma can be interpreted as a proof that $(S_n - m(n))/\sigma_n$ converge in distribution to a normal law as $n \to \infty$. This is not surprising in view of the inversion relation

$$P(S_n = m(n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_n(\theta) d\theta.$$

Further information on this and related subjects can be found, for example, in Feller [2]. Note that it is possible by truncation methods to prove similar results without assuming the existence of moments.

Define $\chi_{k,n}^2$ from $X_{k,n}$ as in Lemma A.1 of the appendix, and let

$$\chi_n^2 \stackrel{\text{def}}{=} \chi_{1,n}^2 + \cdots + \chi_{n,n}^2 \le \sigma_n^2.$$

Lemma 3.1 *Define, for any* $0 < r \le 1$,

$$\delta_n^r \stackrel{\text{def}}{=} \frac{1}{2} \frac{\sigma_n^2}{\beta_n^{(2+r)}}.$$

Let $\chi_n \delta_n \to \infty$ *as* $n \to \infty$. *Then we have the following estimates.*

(*i*) *For any integer* $a \ge 0$,

$$\int_{-\pi}^{\pi} \theta^a \varphi_n(\theta) d\theta = \frac{1}{\sigma_n^{1+a}} \int_{-\infty}^{\infty} u^a e^{-\frac{u^2}{2}} du + O\left(\frac{1}{\sigma_n^{1+a}} \frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\right) + O\left(\frac{1}{\chi_n^2 \delta_n^{1-a}} \exp\left(-\frac{\chi_n^2 \delta_n^2}{5}\right)\right). \quad (3.1)$$

(*ii*) *When* a *is an even number, Equation (3.1) holds for* $1 < r \le 2$, *provided that* $X_{1,n}, \ldots, X_{n,n}$ *admit moments of order* $2+r$ *and choosing*

$$\delta_n = \frac{1}{2} \frac{\sigma_n^2}{\beta_n^{(3)}}.$$

(*iii*) *For any real number* $a \geq 0$,

$$\int_{-\pi}^{\pi} |\theta^a \varphi_n(\theta)| d\theta = O\Big(\frac{1}{\sigma_n^{1+a}}\Big) + O\Big(\frac{1}{\chi_n^2 \delta_n^{1-a}} \exp\Big(-\frac{\chi_n^2 \delta_n^2}{5}\Big)\Big). \qquad (3.2)$$

Proof We only prove (*i*) and (*ii*), since (*iii*) follows trivially from the computations. Define,

$$\begin{aligned} I_n &\stackrel{\text{def}}{=} \int_{-\pi}^{\pi} \theta^a \varphi_n(\theta) d\theta - \int_{-\infty}^{\infty} \theta^a e^{-\frac{\sigma_n^2 \theta^2}{2}} d\theta \\ &= \int_{-\delta_n}^{\delta_n} \theta^a \Big(\varphi_n(\theta) - e^{-\frac{\sigma_n^2 \theta^2}{2}}\Big) d\theta \\ &\quad - \int_{|\theta| \geq \delta_n} \theta^a e^{-\frac{\sigma_n^2 \theta^2}{2}} d\theta + \int_{\delta_n \leq |\theta| \leq \pi} \theta^a \varphi_n(\theta) d\theta, \end{aligned}$$

noting that $\delta_n \leq 1/2$. It is known that

$$\int_{|\theta| \geq \delta_n} \theta^a e^{-\frac{\sigma_n^2 \theta^2}{2}} d\theta \sim \frac{2}{\sigma_n^2 \delta_n^{1-a}} e^{-\frac{\sigma_n^2 \delta_n^2}{2}}.$$

Applying Lemma A.1 to φ_n, we get

$$\Big|\int_{\delta_n \leq |\theta| \leq \pi} \theta^a \varphi_n(\theta) d\theta\Big| \leq \int_{|\theta| \geq \delta_n} |\theta|^a e^{-\frac{\chi_n^2 \theta^2}{5}} d\theta = O\Big(\frac{1}{\chi_n^2 \delta_n^{1-a}} e^{-\frac{\chi_n^2 \delta_n^2}{5}}\Big).$$

Finally,

$$\begin{aligned} |I_n| &\leq \int_{-\delta_n}^{\delta_n} |\theta|^a \Big|\varphi_n(\theta) - e^{-\frac{\sigma_n^2 \theta^2}{2}}\Big| d\theta \\ &\quad + O\Big(\frac{1}{\sigma_n^2 \delta_n^{1-a}} e^{-\frac{\sigma_n^2 \delta_n^2}{2}}\Big) + O\Big(\frac{1}{\chi_n^2 \delta_n^{1-a}} e^{-\frac{\chi_n^2 \delta_n^2}{5}}\Big). \end{aligned} \qquad (3.3)$$

We proceed now to estimate the above integral, so that implicitly $|\theta| \leq \delta_n$. The derivation relies on the following simple inequality, valid for all complex numbers $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$:

$$|x_1 \cdots x_n - y_1 \cdots y_n| \leq \sum_{k=1}^{n} |x_1 \cdots x_{k-1}| |x_k - y_k| |y_{k+1} \cdots y_n|, \qquad (3.4)$$

which will be used with $x_k = \varphi_{k,n}(\theta)$ and $y_k = \exp(-\sigma_{k,n}^2 \theta^2/2)$.

The characteristic function $\varphi_{k,n}$ of the random variable $X_{k,n}$ satisfies (see for example Loève [7])

$$\Big|\varphi_{k,n}(\theta) - 1 + \sigma_{k,n}^2 \frac{\theta^2}{2}\Big| \leq \beta_{k,n}^{(2+r)} \frac{|\theta|^{2+r}}{2}. \qquad (3.5)$$

This implies that

$$\begin{aligned}\left|\varphi_{k,n}(\theta)-e^{-\frac{\sigma_{k,n}^2\theta^2}{2}}\right| &\le \left|\varphi_{k,n}(\theta)-1+\sigma_{k,n}^2\frac{\theta^2}{2}\right|+\left|e^{-\frac{\sigma_{k,n}^2\theta^2}{2}}-1+\sigma_{k,n}^2\frac{\theta^2}{2}\right|\\ &\le \beta_{k,n}^{(2+r)}\frac{|\theta|^{2+r}}{2}+\sigma_{k,n}^{2+r}\frac{|\theta|^{2+r}}{2} \le \beta_{k,n}^{(2+r)}|\theta|^{2+r}. \qquad (3.6)\end{aligned}$$

We have to find an upper bound for $|\varphi_{k,n}|$, assuming first $\sigma_{k,n}\delta_n \le 1$, so that

$$\begin{aligned}|\varphi_{k,n}(\theta)| &\le 1-\sigma_{k,n}^2\frac{\theta^2}{2}+\beta_{k,n}^{(2+r)}\frac{|\theta|^{2+r}}{2}\\ &\le \exp(-\sigma_{k,n}^2+\beta_{k,n}^{(2+r)}\delta_n^r)\frac{\theta^2}{2} \stackrel{\text{def}}{=} \psi_{k,n}(\theta). \qquad (3.7)\end{aligned}$$

Assume now that $\sigma_{k,n}\delta_n \ge 1$. Then,

$$-\sigma_{k,n}^2+\beta_{k,n}^{(2+r)}\delta_n^r \ge -\sigma_{k,n}^2+\sigma_{k,n}^{2+r}\delta_n^r \ge 0,$$

and $|\varphi_{k,n}(\theta)| \le 1 \le \psi_{k,n}(\theta)$.

Since, from the statement of the lemma, $\delta_n\sigma_n \to \infty$, we can choose n to ensure $\delta_n \ge 2/\sigma_n$. Thus,

$$\psi_{k,n}(\theta) \ge \exp\Big(-\sigma_{k,n}^2\frac{\theta^2}{2}\Big) \ge \exp\Big(-\sigma_n^2\frac{\theta^2}{8}\Big).$$

Using (3.4), (3.6) and (3.7), we find

$$\begin{aligned}\left|\varphi_n(\theta)-e^{-\frac{\sigma_n^2\theta^2}{2}}\right| &\le \beta_n^{(2+r)}|\theta|^{2+r}\exp\Big(-\sigma_n^2+\beta_n^{(2+r)}\delta_n^r+\frac{\sigma_n^2}{4})\Big)\frac{\theta^2}{2}\\ &\le \beta_n^{(2+r)}|\theta|^{2+r}\exp\Big(-\sigma_n^2\frac{\theta^2}{8}\Big), \qquad (3.8)\end{aligned}$$

and (3.1) follows, since the integral in (3.3) is bounded by

$$\begin{aligned}\int_{-\delta_n}^{\delta_n}|\theta|^a\left|\varphi_n(\theta)-e^{-\frac{\sigma_n^2\theta^2}{2}}\right|d\theta &\le \beta_n^{(2+r)}\int_{-\infty}^{\infty}|\theta|^{2+r+a}\exp\Big(-\sigma_n^2\frac{\theta^2}{8}\Big)d\theta\\ &= O\Big(\frac{1}{\sigma_n^{1+a}}\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\Big).\end{aligned}$$

The proof of assertion *(ii)* of the lemma is similar, although the computations be more involved. In this case, since a is even , I_n can be rewritten as

$$I_n \stackrel{\text{def}}{=} \int_{-\pi}^{\pi}\theta^a\varphi_n(\theta)d\theta-\int_{-\infty}^{\infty}\theta^a\Big(1-i\bar\beta_n^{(3)}\frac{\theta^3}{6}\Big)e^{-\frac{\sigma_n^2\theta^2}{2}}d\theta,$$

where $\bar\beta_{k,n}^{(3)} \stackrel{\text{def}}{=} \mathbf{E}[X_{k,n}-m_{k,n}]^3$ and $\bar\beta_n^{(3)} \stackrel{\text{def}}{=} \mathbf{E}[X_n-m(n)]^3 = \bar\beta_{1,n}^{(3)}+\cdots+\bar\beta_{n,n}^{(3)}$. To find a bound for $|I_n|$, we have to estimate

$$\left|\varphi_n(\theta) - \left(1 - i\bar\beta_n^{(3)}\frac{\theta^3}{6}\right)e^{-\frac{\sigma_n^2\theta^2}{2}}\right| \tag{3.9}$$
$$\leq \quad \left|\varphi_n(\theta) - e^{-\frac{\sigma_n^2\theta^2}{2} - i\frac{\bar\beta_n^{(3)}\theta^3}{6}}\right| + \left|e^{-i\frac{\bar\beta_n^{(3)}\theta^3}{6}} - 1 + i\bar\beta_n^{(3)}\frac{\theta^3}{6}\right|e^{-\frac{\sigma_n^2\theta^2}{2}}.$$

The second part of the r.h.s. of (3.9) is trivially less than

$$\left|\bar\beta_n^{(3)}\frac{\theta^3}{6}\right|^{1+\frac{r}{3}} e^{-\frac{\sigma_n^2\theta^2}{2}} \leq \beta_n^{(3+r)}\frac{|\theta|^{3+r}}{6}e^{-\frac{\sigma_n^2\theta^2}{2}},$$

and the first one is evaluated as above with (3.4) and

$$\psi_{k,n}(\theta) = \exp(-\sigma_{k,n}^2 + \beta_{k,n}^{(3)}\delta_n)\frac{\theta^2}{2}.$$

■

We are now in a position to state the main result of this section, which involves a Lyapounov's condition, popular in the Central Limit Problem.

Theorem 3.2 *Let $s \leq r$ be two real numbers such that $0 < r \leq 2$ and $0 < s \leq 1$. Assume that $\sigma_n = O(\chi_n)$, that $\beta_n^{(2+r)}$ exists and $\beta_n^{(2+r)}/\sigma_n^{2+r} \to 0$ as $n \to \infty$. Then the following asymptotic expansions hold.*

(*i*)

$$Z_n = \frac{F_n(\lambda_n)}{\sigma_n\sqrt{2\pi}}\left[1 + O\left(\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\right)\right]. \tag{3.10}$$

(*ii*) *For any $j \leq n$,*

$$\mathbf{E}\,Q_{j,n} = \mathbf{E}\,X_{j,n} + O\left(\frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}}\right) + O\left(\frac{\beta_n^{(2+s)}\sigma_{j,n}^2 + \beta_{j,n}^{(2+s)}\sigma_n^2}{\sigma_n^{3+s}}\right). \tag{3.11}$$

(*iii*) *For any $\ell \leq n$, if $[\sum_{j=1}^{\ell} m_{j,n}]/\sigma_n \to 0$,*

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) = \prod_{k=1}^{\ell} P(X_{k,n} = q_k)\left[1 + O(\varepsilon_n)\right], \tag{3.12}$$

$$\varepsilon_n = \frac{\beta_n^{(2+r)}}{\sigma_n^{2+r}} + \frac{\beta_n^{(2+s)}\sum_{j=1}^{\ell} m_{j,n}}{\sigma_n^{3+s}} + \frac{[\sum_{j=1}^{\ell} m_{j,n}]^2}{\sigma_n^2}.$$

Proof Equation (3.10) is a simple application of Lemma 3.1 to (2.6), since

$$\begin{aligned} Z_n &= \frac{F_n(\lambda_n)}{2\pi}\int_{-\pi}^{\pi}\prod_{k=1}^{n}\varphi_{k,n}(\theta)d\theta \\ &= \frac{F_n(\lambda_n)}{2\pi}\left[\frac{\sqrt{2\pi}}{\sigma_n}+O\Big(\frac{\beta_n^{(2+r)}}{\sigma_n^{3+r}}\Big)+O\Big(\frac{1}{\chi_n^2\delta_n}\exp\big(-\frac{\chi_n^2\delta_n^2}{5}\big)\Big)\right]. \end{aligned}$$

Like in (3.5), we can write

$$\varphi'_{j,n}(\theta) = -\sigma_{j,n}^2\theta + O(\beta_{j,n}^{(2+s)})\theta^{1+s},$$

and (2.7) reads:

$$\begin{aligned} \mathbf{E}\,Q_{j,n} &= m_{j,n} + \frac{F_n(\lambda_n)}{2\pi Z_n}\Big[i\sigma_{j,n}^2\int_{-\pi}^{\pi}\theta\prod_{k\neq j}^{n}\varphi_{k,n}(\theta)d\theta \\ &\quad + O(\beta_{j,n}^{(2+s)})\int_{-\pi}^{\pi}\theta^{1+s}\prod_{k\neq j}^{n}\varphi_{k,n}(\theta)d\theta\Big] \\ &= m_{j,n} + \frac{F_n(\lambda_n)}{2\pi Z_n}\left[O\Big(\frac{\beta_n^{(2+s)}\sigma_{j,n}^2}{\sigma_n^{4+s}}\Big)+O\Big(\frac{\beta_{j,n}^{(2+s)}}{\sigma_n^{2+s}}\Big)\right], \end{aligned}$$

and (3.11) follows. Equation (3.12) is obtained in the same way from (2.8) and

$$e^{i\sum_{j=1}^{\ell}(q_j-m_{j,n})\theta} = 1 - i\Big[\sum_{j=1}^{\ell} q_j - m_{j,n}\Big]\theta + O\Big(\Big[\sum_{j=1}^{\ell} q_j - m_{j,n}\Big]^2\Big)\theta^2.$$

■

4 Asymptotic expansion with heavily loaded queues

In this section, we analyze the network $\mathcal{B}_n$ when some queues become heavily loaded. This, in particular, implies that the Lyapounov condition $\beta_n^{(2+r)}/\sigma_n^{2+r} \to 0$ is no more valid. In fact, after proper scaling, $S_n - m(n)$ will be shown to converge in distribution to a random variable having a gamma distribution, under the broad assumption that the first singularities of the relevant generating functions are algebraic.

Lemma 4.1 *Assume that we can write for some fixed $\xi \geq 1$*

$$\varphi_n(\theta) = \omega_n^{\xi}(\theta)\widehat{\varphi}_n(\theta),$$

where $\widehat{\varphi}_n$ is a characteristic function and

$$\omega_n(\theta) \overset{def}{=} \frac{(1-\rho_n)e^{-i\alpha_n\theta}}{1-\rho_n e^{i\theta}},$$

with $0 \le \rho_n < 1$ and $\alpha_n \overset{def}{=} \rho_n/(1-\rho_n)$. We fix r, $0 < r \le 1$. Let $\hat{\sigma}_n$, $\hat{\beta}_n^{(2+r)}$, $\hat{\chi}_n$ and $\hat{\delta}_n$ be quantities having the same meaning as in Lemma 3.1, but related to $\widehat{\varphi}_n(\theta)$.

Let $\hat{\sigma}_n/\alpha_n \to 0$ and $\hat{\delta}_n\hat{\chi}_n \to \infty$ as $n \to \infty$. Then the following estimates hold:

$$\begin{aligned}\alpha_n \int_{-\pi}^{\pi} \varphi_n(\theta)d\theta &= \int_{-\infty}^{\infty} \frac{e^{-i\xi u}}{(1-iu)^{\xi}} \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)du \qquad (4.1)\\ &\quad + O\Big(\frac{1}{\alpha_n} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{2+r} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi-1}\Big)\\ &\quad + O\Big(\frac{e^{-\frac{\hat{\chi}_n^2\hat{\delta}_n^2}{5}}}{\hat{\chi}_n^2\hat{\delta}_n^{\xi+1}\alpha_n^{\xi-1}}\Big),\\ \alpha_n \int_{-\pi}^{\pi} |\theta\varphi_n(\theta)|d\theta &= O\Big(\frac{1}{\alpha_n} + \frac{1}{\hat{\sigma}_n}\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi-1}\Big) + O\Big(\frac{e^{-\frac{\hat{\chi}_n^2\hat{\delta}_n^2}{5}}}{\hat{\chi}_n^2\hat{\delta}_n^{\xi}\alpha_n^{\xi-1}}\Big). \qquad (4.2)\end{aligned}$$

Proof The proof of this lemma is similar to the proof of Lemma 3.1 and is only sketched here. Define

$$\omega(u) \overset{def}{=} \frac{e^{-iu}}{1-iu},$$

and

$$\begin{aligned}I_n &\overset{def}{=} \alpha_n \int_{-\pi}^{\pi} \omega_n^{\xi}(\theta)\widehat{\varphi}_n(\theta)d\theta - \int_{-\infty}^{\infty} \omega^{\xi}(u)\exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)du\\ &= \int_{-\pi\alpha_n}^{\pi\alpha_n} \omega_n^{\xi}(u/\alpha_n)\Big[\widehat{\varphi}_n(u/\alpha_n) - \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)\Big]du\\ &\quad + \int_{-\pi\alpha_n}^{\pi\alpha_n} \Big[\omega_n^{\xi}(u/\alpha_n) - \omega^{\xi}(u)\Big]\exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)du\\ &\quad - \int_{|u|\ge\pi\alpha_n} \omega^{\xi}(u)\exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)du. \qquad (4.3)\end{aligned}$$

The evaluation of these integrals depends on the following straightforward estimations, valid for $|u| < \pi\alpha_n$:

$$\begin{aligned}|\omega_n^{\xi}(u/\alpha_n)| &= O\Big(\frac{1}{(1+u^2)^{\xi/2}}\Big)\\ |\omega_n^{\xi}(u/\alpha_n) - \omega^{\xi}(u)| &= O\Big(\frac{1}{\alpha_n}\frac{u^2}{(1+u^2)^{\xi}}\Big),\end{aligned}$$

and from (3.8), which reads for $|u| < \alpha_n \hat{\delta}_n$:

$$\left|\widehat{\varphi}_n(u/\alpha_n) - \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\Big)\right| = O\Big(\frac{\hat{\beta}_n^{(2+r)}}{\alpha_n^{2+r}}\Big) u^{2+r} \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{8}\Big).$$

Moreover, we use the following approximation, valid for $a, b > 0$ and for small z:

$$J(a,b,z) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \frac{|u|^a}{(1+u^2)^b} e^{-z^2u^2} du = O(1) + O(z^{2b-a-1}).$$

These relations, together with (4.3), yield:

$$\begin{aligned} I_n &= O\Big(\frac{1}{\alpha_n}\frac{\hat{\beta}_n^{(2+r)}}{\alpha_n^{2+r}}\Big) J\Big(2+r, \frac{\xi}{2}, \frac{\hat{\sigma}_n}{\alpha_n}\Big) + O\Big(\frac{1}{\alpha_n^2}\Big) J\Big(2, \xi, \frac{\hat{\sigma}_n}{\alpha_n}\Big) \\ &\quad + O\Big(\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi-1}\Big) \int_{v \geq \pi\hat{\sigma}_n} v^{-\xi} e^{-\frac{v^2}{2}} dv \\ &\quad + O\Big(\Big(\frac{\hat{\chi}_n}{\alpha_n}\Big)^{\xi-1}\Big) \int_{v \geq \delta_n \hat{\chi}_n} v^{-\xi} e^{-\frac{v^2}{2}} dv, \end{aligned}$$

and (4.1) follows. Relation (4.2) is obtained from the estimation

$$\begin{aligned} \int_{-\delta_n\alpha_n}^{\delta_n\alpha_n} \left| u\omega_n^{\xi}(u/\alpha_n)\varphi_n(u/\alpha_n)\right| du &\leq \int_{-\infty}^{\infty} \frac{|u|}{(1+u^2)^{\xi/2}} \exp\Big(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{4}\Big) du \\ &= O(1) + O\Big(\Big(\frac{\hat{\sigma}_n}{\alpha_n}\Big)^{\xi-2}\Big). \end{aligned}$$

■

The estimates of Lemma 4.1 allow to establish the main result of this section, with the same notation.

Theorem 4.2 *Assume that the queues can be split into two categories:*

(*i*) *a set $\mathcal{S}_n$ of "saturated" queues, for which*

$$\varphi_{k,n}(\theta) = \omega_n^{\xi_{k,n}}(\theta)\psi_{k,n}(\theta),$$

with $\psi_{k,n}(\theta) = O(1)$ and $\psi'_{k,n}(\theta) = O(\theta)$, uniformly in k and n;

(*ii*) *"normal" queues, with total characteristic function $\widehat{\varphi}_n(\theta)$.*

Define $\xi_n = \sum_{k\in\mathcal{S}_n} \xi_{k,n}$ and let r be a real number, $0 < r \leq 1$.

Assume $\hat{\sigma}_n/\alpha_n \to 0$, $\hat{\beta}_n^{(2+r)}/\hat{\sigma}_n^{2+r} \to 0$ and $\hat{\sigma}_n = O(\hat{\chi}_n)$ as $n \to \infty$. Then the following expansions hold.

(*i*)

$$Z_n = \frac{F_n(\lambda_n)}{\alpha_n} \frac{\xi_n^{\xi_n - 1} e^{-\xi_n}}{\Gamma(\xi_n)} [1 + O(\varepsilon_{1,n})], \tag{4.4}$$

with

$$\varepsilon_{1,n} = \left(\frac{\hat{\sigma}_n}{\alpha_n}\right)^2 + \frac{1}{\alpha_n} + \frac{\hat{\beta}_n^{(2+r)}}{\hat{\sigma}_n^{2+r}} \left(\frac{\hat{\sigma}_n}{\alpha_n}\right)^{\xi_n - 1};$$

(*ii*) *for any* $j \le n$,

$$\mathbf{E}\, Q_{j,n} = \mathbf{E}\, X_{j,n}[1 + O(\varepsilon_{1,n} + \varepsilon_{2,n})], \tag{4.5}$$

with

$$\varepsilon_{2,n} = \begin{cases} \dfrac{\sigma_{j,n}^2}{\alpha_n} + \dfrac{\sigma_{j,n}^2}{\hat{\sigma}_n}\left(\dfrac{\hat{\sigma}_n}{\alpha_n}\right)^{\xi_n - 1} & \text{if } j \notin S_n, \\ \dfrac{1}{\alpha_n} + \dfrac{1}{\hat{\sigma}_n}\left(\dfrac{\hat{\sigma}_n}{\alpha_n}\right)^{\xi_n - 1} & \text{otherwise;} \end{cases}$$

(*iii*) *for any* $\ell \le n$, *if* $S_n \cap [1, j] = \emptyset$,

$$P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_\ell) = \prod_{k=1}^{n} P(X_{k,n} = q_k)\Big[1 + O(\varepsilon_{1,n} + \varepsilon_{3,n})\Big], \tag{4.6}$$

with

$$\varepsilon_{3,n} = \sum_{j=1}^{\ell} m_{j,n}\left(\frac{1}{\alpha_n} + \frac{1}{\hat{\sigma}_n}\left(\frac{\hat{\sigma}_n}{\alpha_n}\right)^{\xi_n - 1}\right).$$

Proof The proof is essentially identical to the proof of Theorem 3.2. The first term of (4.1) is evaluated using Parseval's identity and classical complex analysis tools, giving

$$\int_{-\infty}^{\infty} \omega_n^{\xi}(u) \exp\left(-\frac{\hat{\sigma}_n^2}{\alpha_n^2}\frac{u^2}{2}\right) du = \frac{\xi_n^{\xi_n - 1} e^{-\xi_n}}{\Gamma(\xi_n)}\left[1 + O\left(\frac{\hat{\sigma}_n^2}{\alpha_n^2}\right)\right].$$

■

5 The choice of a good scaling

One thing that has been left out in the previous sections is the determination of "interesting" values for $m(n)$ or, equivalently, for λ_n. The analysis in [4, 5, 6, 1, 8] was done assuming that the number of customers in the system should be proportional to the number of stations. Indeed, we shall see that this is not always the suitable scaling. In order to exhibit meaningful behaviours, it will be convenient to make additional assumptions on the model. However, it should be pointed out that the results of Sections 3 and 4 are valid in a more general setting, which would allow to carry out special cases via the same methods.

The queues are partitioned as follows:

$$\begin{aligned}
\mathcal{F}_n &\stackrel{\text{def}}{=} \{0 \le k \le n \,:\, \liminf_{q \to \infty} \mu_{k,n}(q) < +\infty\},\\
\mathcal{I}_n &\stackrel{\text{def}}{=} \{0 \le k \le n \,:\, \lim_{q \to \infty} \mu_{k,n}(q) = +\infty\},
\end{aligned}$$

and we define

$$\begin{aligned}
\mu_{k,n} &\stackrel{\text{def}}{=} \begin{cases} \liminf_{q\to\infty} \mu_{k,n}(q) & \text{if } k \in \mathcal{F}_n \\ \mu_{k,n}(1) & \text{if } k \in \mathcal{I}_n \end{cases},\\
\rho_{k,n} &\stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n}}{\mu_{k,n}},\\
\rho_n &\stackrel{\text{def}}{=} \sup_{k \in \mathcal{F}_n} \rho_{k,n},\\
\frac{1}{\bar{\mu}_n(1)} &\stackrel{\text{def}}{=} \sum_{k=1}^{n} \frac{\pi_{k,n}}{\mu_{k,n}(1)},
\end{aligned}$$

The analysis is done under the following assumption on the service disciplines.

Assumption A1 *There exist sequences* $R(q)$ *and* $T(q)$ *such that*

$$\lim_{q\to\infty} R(q) = 1,$$

$$\lim_{q\to\infty} T(q) = +\infty,$$

and, for any $q > 0$,

$$\begin{aligned}
\frac{\mu_{k,n}(q)}{\mu_{k,n}} &\ge R(q) \quad \textit{for } k \in \mathcal{F}_n,\\
\frac{\mu_{k,n}(q)}{\mu_{k,n}} &\ge T(q) \quad \textit{for } k \in \mathcal{I}_n.
\end{aligned}$$

This assumption holds in particular for any mixing of infinite-server and multiple-server queues, the number of servers remaining bounded. Moreover, for the sake of simplicity, we will assume

Assumption A2 $\bar{\mu}_n(1) = \Omega(1)$.

This ensures that the mean service times are bounded away from zero and from infinity. When this assumption is not satisfied, the results of this section still hold, with λ_n replaced by $\lambda_n/\bar{\mu}_n(1)$ in the error terms.

The first theorem deals with scalings, where no queue of the network $\mathcal{B}_n$ becomes saturated.

Theorem 5.1 *Assume that* **A1** *and* **A2** *hold and that there exists constants* A *and* B *such that*

$$\lim_{n\to\infty} \lambda_n = +\infty, \tag{5.1}$$

$$\rho_{k,n} \le B < 1, \text{ for all } k \in \mathcal{F}_n, \tag{5.2}$$

$$\rho_{k,n} \le A, \text{ for all } k \in \mathcal{I}_n. \tag{5.3}$$

Then the following expansions hold, for all $j \le n$.

$$Z_n = \frac{F_n(\lambda_n)}{\sigma_n\sqrt{2\pi}}\left[1 + O\left(\frac{1}{\lambda_n}\right)\right], \tag{5.4}$$

$$\mathbf{E}\, Q_{j,n} = \mathbf{E}\, X_{j,n}\left[1 + O\left(\frac{1}{\lambda_n}\right)\right], \tag{5.5}$$

$$\begin{aligned} &P(Q_{1,n} = q_1, \ldots, Q_{\ell,n} = q_j) \\ &\quad = P(X_{1,n} = q_1, \ldots, X_{\ell,n} = q_j)\left[1 + O\left(\frac{1}{\lambda_n}\right)\right]. \end{aligned} \tag{5.6}$$

Moreover, $m(n) = \Omega(\lambda_n)$.

Proof This theorem is a simple rewriting of Theorem 3.2, using **A1** and (5.1)–(5.3). For a queue $k \in \mathcal{F}_n$, and for all $a \in \mathbb{N}$, we have

$$\sum_{q=0}^{\infty} q^a \frac{(\lambda_n \pi_{k,n})^q}{\mu_{k,n}(1)\cdots\mu_{k,n}(q)} \le \sum_{q=0}^{\infty} \frac{q^a B^q}{R(1)\cdots R(q)} < +\infty.$$

In particular, $f_{k,n}(\lambda_n\pi_{k,n}) = \Omega(1)$ and

$$m_{k,n} = \frac{\lambda_n\pi_{k,n}}{\mu_{k,n}(1) f_{k,n}(\lambda_n\pi_{k,n})} \sum_{q=1}^{\infty} q \frac{(\lambda_n\pi_{k,n})^{q-1}}{\mu_{k,n}(2)\cdots\mu_{k,n}(q)} = \Omega\left(\frac{\lambda_n\pi_{k,n}}{\mu_{k,n}(1)}\right).$$

Similarly, for any $r \in \mathbb{N}$,

$$\beta_{k,n}^{(r)} = \Omega\left(\frac{\lambda_n\pi_{k,n}}{\mu_{k,n}(1)}\right),$$

and the results of Theorem 3.2 apply. The remaining computations are done in a similar fashion for all types of queues. ∎

Theorem 5.1 can be easily generalized to the situation where some queues of $\mathcal{I}_n$ become saturated, in which case (5.3) fails to be satisfied.

Corollary 5.2 *Let* **A1**, **A2**, *(5.1) and (5.2) be satisfied and assume that the queues for which (5.3) does not hold are infinite-server queues. Then the conclusions of Theorem 5.1 remain valid, except (5.6), which holds only for queues satisfying (5.2) or (5.3).*

Proof This proof relies on the fact that the characteristic function of the number of clients X in an infinite server queue with parameter ρ can be written as

$$\mathbf{E}\, e^{i\theta X} = \exp\Big(\rho(e^{i\theta} - 1)\Big) = \Big[\exp\Big(\frac{\rho}{\lfloor\rho\rfloor}(e^{i\theta} - 1)\Big)\Big]^{\lfloor\rho\rfloor},$$

which means that a saturated infinite server queue can be replaced by several non-saturated infinite-server queues without changing the distribution of S_n. With this substitution, the proof Theorem 5.1 applies without change. ■

To handle the case of saturated queues in $\mathcal{F}_n$, we have to introduce the subset of $\mathcal{F}_n$

$$\mathcal{S}_n \stackrel{\text{def}}{=} \{k \in \mathcal{F}_n \ : \ \rho_{k,n} \doteq \rho_n\},$$

and to assume:

Assumption A3 *The cardinal of* $\mathcal{S}_n$ *is uniformly bounded and, for all* $k \in \mathcal{S}_n$, $f_{k,n}(z)$ *has its first singularity at* $z = \mu_{k,n}$, *which is algebraic of order* $\xi_{k,n}$. *Moreover, for some* $\xi > 0$,

$$\xi_n \stackrel{\text{def}}{=} \sum_{k \in \mathcal{S}_n} \xi_{k,n} \in [1, \xi].$$

While this assumption covers a wide range of known queues, other kind of singularities could be handled via the same method.

The next theorem illustrates what happens when $\rho_n \to 1$.

Theorem 5.3 *Assume that* **A1**, **A2** *and* **A3** *hold and that there exists constants* A *and* B *such that*

$$\lim_{n\to\infty} \rho_n = 1, \tag{5.7}$$

$$\rho_{k,n} \le B < 1, \text{ for all } k \in \mathcal{F}_n \setminus \mathcal{S}_n, \tag{5.8}$$

$$\rho_{k,n} \le A, \text{ for all } k \in \mathcal{I}_n. \tag{5.9}$$

(*i*) *If* $\lim_{n\to\infty}(1-\rho_n)^2\lambda_n = +\infty$, *then the following expansions hold for any* $j \le n$:

$$\begin{aligned} Z_n &= \frac{F_n(\lambda_n)}{\sigma_n\sqrt{2\pi}}\Big[1 + O\Big(\frac{1}{\lambda_n} + \frac{1}{\lambda_n^2(1-\rho_n)^4}\Big)\Big], \\ \mathbf{E}\, Q_{j,n} &= \mathbf{E}\, X_{j,n}\Big[1 + O\Big(\frac{1}{\lambda_n} + \frac{1}{\lambda_n^2(1-\rho_n)^4}\Big)\Big], \ \textit{if } j \notin \mathcal{S}_n, \end{aligned}$$

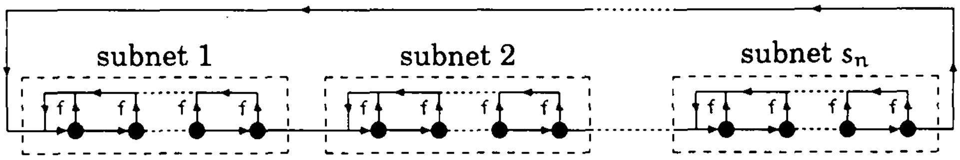

Figure 1: a compound network of tandem queues

$$\begin{aligned}
\mathbf{E}\, Q_{j,n} &= \mathbf{E}\, X_{j,n}\Big[1 + O\Big(\frac{1}{\lambda_n(1-\rho_n)^2}\Big)\Big], \ \textit{if } j \in \mathcal{S}_n, \\
P(Q_{1,n} &= q_1, \ldots, Q_{\ell,n} = q_j) \\
&= \prod_{k=1}^{n} P(X_{k,n} = q_k)\Big[1 + O\Big(\frac{1}{\lambda_n} + \frac{1}{\lambda_n^2(1-\rho_n)^4}\Big)\Big], \\
&\qquad\qquad \textit{if } [1, j] \cap \mathcal{S}_n = \emptyset.
\end{aligned}$$

(*ii*) *if* $\lim_{n\to\infty}(1-\rho_n)^2\lambda_n = 0$, *then the expansions of Theorem 4.2 hold with* $r = 1$ *and*

$$\hat{\beta}_n^{(3)} = \Omega(\lambda_n), \quad \hat{\sigma}_n^2 = \Omega(\lambda_n),$$
$$\alpha_n = \Omega\Big(\frac{1}{1-\rho_n}\Big), \qquad m(n) = \Omega\Big(\frac{1}{1-\rho_n}\Big) + \Omega(\lambda_n).$$

The proof of this theorem is straightforward and is therefore omitted. There is also an obvious equivalent of Corollary 5.2, (when (5.9) does not hold) which will not be explicitly stated.

As pointed out at the beginning of this section, there are cases of interest with $\lambda_n = o(n)$ and $m(n) = o(n)$. This will be illustrated in the next example.

Example Consider a closed network consisting of s_n subnetworks of $M/M/1$ queues having each a unique entry point, in which a fixed number m of tasks circulate. The queues are subject to failures, taking place with some probability f. When a failure occurs, the task returns to the entry point of its current subnetwork. Tasks visit the various subnetworks according to some probability matrix.

This model exhibits tight bottlenecks, when the number and the size of the subnetworks grow. This fact, for the sake of simplicity, will be illustrated on a very simple topology, presented in Figure 1: all subnetworks are associated in tandem, and each of them consists itself of ℓ_n queues in tandem, with unit processing rates.

Here, the invariant measure of the routing matrix has the form

$$(\pi_1, \ldots, \pi_{\ell_n}; \pi_1, \ldots; \ldots, \pi_{\ell_n}),$$

where π_k is the invariant probability associated to the k-th queue of any subnetwork. A straightforward computation, using symmetry properties, yields

$$\pi_k = \frac{1}{s_n}\frac{f(1-f)^k}{1-(1-f)^{\ell_n}}.$$

Therefore, the s_n entry points of the subnetworks will act as bottlenecks for the system, since the size of the queues remain uniformly bounded if, and only if, $\lambda_n \pi_1 < B < 1$. It suffices to take

$$\lambda_n \leq \frac{s_n B}{f(1-f)}.$$

Thus, from Theorem 5.1, it follows that

$$m(n) = \Omega(\lambda_n) = \Omega(s_n).$$

Thus, since $n = \ell_n s_n$, the function $m(n)$ has an arbitrary asymptotic behaviour.

6 Application to a service vehicle network

Consider a fleet of vehicles serving an area consisting of n stations forming a fully connected graph. These vehicles are used to transport goods or passengers. Vehicles wait at stations until they receive a request, in which case they go to an other station. The routing among stations is done according to some routing matrix P_n. When a request arrives to an empty station, it is immediately lost. The request arrivals form a Poisson stream at each queue.

We model this system as follows: for all $0 \leq k \leq n$, station k is represented as a single-server queue with service rate $\mu_{k,n}$ which is equal to the arrival rate at station k, since arrivals are lost when the station is empty. When a vehicle leaves station k, it chooses its destination according to the Markovian routing matrix $P_n = (p_{k\ell,n})$. The duration of the journey between two stations k and ℓ is represented by an infinite server queue placed on the edge between them. The service rate of this queue when there are q vehicles traveling between k and ℓ is $\mu_{k\ell,n} q$. Note that, contrary to the convention used throughout this paper, the total number of *queues* is $n^2 + n$. Let $(\pi_{1,n}, \ldots, \pi_{n,n})$ be the invariant measure of P_n, defined as in (2.1). Then, with obvious notation, for all $k, \ell \in [1, n]$, for all $\theta \in [-\pi, \pi]$,

$$\rho_{k,n} \stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n}}{2\mu_{k,n}},$$

$$
\begin{aligned}
m_{k,n} &\stackrel{\text{def}}{=} \frac{\rho_{k,n}}{1-\rho_{k,n}},\\
\varphi_{k,n}(\theta) &\stackrel{\text{def}}{=} \frac{(1-\rho_{k,n})e^{-im_{k,n}\theta}}{1-\rho_{k,n}e^{i\theta}},\\
\rho_{k\ell,n} &\stackrel{\text{def}}{=} \frac{\lambda_n \pi_{k,n} p_{k\ell,n}}{2\mu_{k\ell,n}},\\
m_{k\ell} &\stackrel{\text{def}}{=} \rho_{k\ell,n},\\
\varphi_{k\ell,n}(\theta) &\stackrel{\text{def}}{=} \exp\Big(\rho_{k\ell,n}(e^{i\theta}-1-i\theta)\Big),\\
\varphi_n(\theta) &\stackrel{\text{def}}{=} \exp\left[(e^{i\theta}-1-i\theta)\sum_{k,\ell=1}^{n}\rho_{k\ell,n}\right]\prod_{k=1}^{n}\frac{(1-\rho_{k,n})e^{-im_{k,n}\theta}}{1-\rho_{k,n}e^{i\theta}},
\end{aligned}
$$

Some questions of interest come to mind:

(*i*) what is the maximal efficiency that can be expected from this system?

(*ii*) how many vehicles should be provided?

Among all performance measures which can be derived from Theorems 5.1 and 5.3, we define the *loss probability* as

$$
\mathcal{P}_{\text{loss}}(n) \stackrel{\text{def}}{=} \frac{\sum_{k=1}^{n}\mu_{k,n}\,P(Q_{k,n}=0)}{\sum_{k=1}^{n}\mu_{k,n}}
$$

Moreover, we introduce the *mean station time* and the *mean travel time* of a vehicle, respectively defined as:

$$
\begin{aligned}
\bar{\tau}_n^{(s)} &\stackrel{\text{def}}{=} \sum_{k=1}^{n}\frac{\pi_{k,n}}{2\mu_{k,n}},\\
\bar{\tau}_n^{(t)} &\stackrel{\text{def}}{=} \sum_{k,\ell=1}^{n}\frac{\pi_{k,n}p_{k\ell,n}}{2\mu_{k\ell,n}}.
\end{aligned}
$$

and we assume that these quantities are uniformly bounded

$$
\bar{\tau}_n^{(s)} = \Omega(1) \quad \bar{\tau}_n^{(t)} = \Omega(1),
$$

so that **A2** holds. Note that when these assumptions are not fulfilled, the system can be considered as flawed. Define $\mathcal{S}_n$ as in Section 5 and assume that its cardinal is some fixed integer $K \geq 1$. This implies that assumptions **A1** and **A3** hold. Theorems 5.1 and 5.3 can then be rewritten as:

Theorem 6.1 *Assume that (5.8) holds. Recall from Section 5 that*

$$
\rho_n = \lambda_n \sup_{1\leq k\leq n}\frac{\pi_{k,n}}{2\mu_{k,n}} < 1.
$$

(*i*) *if* $\lim_{n\to\infty}(1-\rho_n)^2\lambda_n = +\infty$ *then*

$$\begin{aligned}
\mathcal{P}_{loss}(n) &= \frac{\sum_{k\notin\mathcal{S}_n}\mu_{k,n}(1-\rho_{k,n})}{\sum_{k=1}^{n}\mu_{k,n}}\left[1+O\left(\frac{1}{\lambda_n}+\frac{1}{\lambda_n^2(1-\rho_n^4)}\right)\right],\\
m(n) &= \Omega(\lambda_n);
\end{aligned}$$

(*ii*) *if* $\lim_{n\to\infty}(1-\rho_n)^2\lambda_n = 0$ *then*

$$\begin{aligned}
\mathcal{P}_{loss}(n) &= \frac{\sum_{k\notin\mathcal{S}_n}\mu_{k,n}(1-\rho_{k,n})}{\sum_{k=1}^{n}\mu_{k,n}}\left[1+O\left(\frac{1}{\sqrt{\lambda_n}}\right)\right],\ \textit{if}\ K=1,\\
\mathcal{P}_{loss}(n) &= \frac{\sum_{k\notin\mathcal{S}_n}\mu_{k,n}(1-\rho_{k,n})}{\sum_{k=1}^{n}\mu_{k,n}}\left[1+O(1-\rho_n)\right],\ \textit{if}\ K\geq 2,\\
m(n) &= \Omega\left(\frac{1}{1-\rho_n}\right)+\Omega(\lambda_n).
\end{aligned}$$

Define the minimum loss probability

$$\mathcal{P}^*_{loss}(n) = \inf_{m(n)\in\mathbb{N}}\mathcal{P}_{loss}(n).$$

Since, as $n\to\infty$,

$$\mathcal{P}^*_{loss}(n) \sim 1-\frac{\inf_{1\leq k\leq n}\frac{\mu_{k,n}}{\pi_{k,n}}}{\sum_{k=1}^{n}\mu_{k,n}},$$

it is useless to increase the number of vehicles in the system beyond a quantity of order

$$m(n) = \Omega\left(\inf_{1\leq k\leq n}\frac{\mu_{k,n}}{\pi_{k,n}}\right).$$

Consequently, having a number of vehicle proportional to the number of stations can be a poor choice, especially when some stations are more loaded than the others. These stations act as *bottlenecks* of the system, which can be removed by changing the routing probabilities.

7 General remarks

(*i*) In Section 3, the reader may be tempted to raise questions about possible caracterizations of the original network by means of other families of limit laws. For example, it is easy to choose a proper normalization a_n ensuring that

$$\frac{S_n}{a_n} = \frac{X_{1,n}+\cdots+X_{n,n]}}{a_n}\xrightarrow{w}\mathcal{P}(\lambda),$$

under standard assumptions, where $\mathcal{P}(\lambda)$ denotes a Poisson law of parameter λ. The key point is then to "lift" these results onto the original network, otherwise they would be of little interest. In fact, this might well be the case for the following reason: most of the estimates in Section 3 were in fact obtained by computing deviations of S_n around $\mathbf{E}\, S_n = m(n)$. In the case just quoted (*i.e.* Poisson limit law), there is no natural reason for working "around the mean".

(*ii*) In the case of saturated queues, additional assumptions had to be introduced. The main reason is that, in this case, the standard "uniform asymptotic negligeability" conditions do not hold, which means in particular that we are no more in the classical framework of the Central Limit Theorem.

A Appendix: a bound on periodic characteristic functions

One of the problems arising in the computation of convergence rates in the Central Limit Theorem is to find upper bounds on the modulus of a characteristic function $\varphi(\theta)$ for θ away from 0. One typical property used is usually stated as follows:

there exist $\theta_0 > 0$ *and* $a < 1$ *such that, for all* $|\theta| > \theta_0$, $|\varphi(\theta)| < a$.

It is pointed out in Feller [2] that this condition is usually easy to fulfill in practice, as long as X does not have a lattice distribution. Unfortunately, we are in the lattice case and we must cope with problem of periodicity of φ.

We show therefore in next lemma how a bound on $|\varphi(\theta)|$ can be derived for $|\theta| \le \pi$.

Lemma A.1 *Let* X *be an integer-valued random variable with distribution* $P(X = k) = p_k$, $k \in \mathbb{N}$. *Define*

$$\chi^2 \stackrel{def}{=} \sum_{k=0}^{\infty} \frac{p_{2k} p_{2k+1}}{p_{2k} + p_{2k+1}},$$

where the summands are taken to be zero when $p_{2k} = p_{2k+1} = 0$. *Then, for any* $\theta \in [0, \pi]$, *the characteristic function* φ *of* X *satisfies:*

$$|\varphi(\theta)| \le \exp\Big(-\frac{\chi^2}{5}\theta^2\Big). \tag{A.1}$$

χ^2 is always less than $\min(\operatorname{Var} X, 1/4)$. *If, in addition, there exists a real number ρ such that $p_{k+1} \le \rho p_k$, then*

$$\frac{1-p_0}{(1+\rho)^2} \le \chi^2 \le \frac{\rho(p_0+\rho)}{(1+\rho)^2}. \tag{A.2}$$

Proof We have

$$|\varphi(\theta)| = \Big|\sum_{k=0}^{\infty} p_k e^{ik\theta}\Big| \le \sum_{k=0}^{\infty} \Big|p_{2k} + p_{2k+1} e^{i\theta}\Big|.$$

Moreover,

$$\begin{aligned}
\Big|p_{2k} + p_{2k+1} e^{i\theta}\Big| &= \sqrt{(p_{2k} + p_{2k+1}\cos\theta)^2 + p_{2k+1}^2 \sin^2\theta} \\
&= \sqrt{(p_{2k} + p_{2k+1})^2 - 2p_{2k}p_{2k+1}(1-\cos\theta)} \\
&\le p_{2k} + p_{2k+1} - \frac{p_{2k}p_{2k+1}}{p_{2k}+p_{2k+1}}(1-\cos\theta).
\end{aligned}$$

Hence, for $\theta \in [0, \pi]$,

$$\begin{aligned}
|\varphi(\theta)| &\le 1 - (1-\cos\theta)\sum_{k=0}^{\infty} \frac{p_{2k}p_{2k+1}}{p_{2k}+p_{2k+1}} \\
&\le 1 - \frac{2}{\pi^2}\theta^2\chi^2 \\
&\le \exp\left(-\frac{2\chi^2}{\pi^2}\theta^2\right),
\end{aligned}$$

which gives (A.1). That $\chi^2 \le \operatorname{Var} X$ can be seen by a Taylor expansion of φ in the neighborhood of $\theta = 0$, while the relation $\chi^2 \le 1/4$ follows from the trivial inequality

$$\frac{p_{2k}p_{2k+1}}{p_{2k}+p_{2k+1}} \le \frac{p_{2k}+p_{2k+1}}{4}.$$

To prove (A.2), remark that, when $p_{k+1} \le \rho p_k$,

$$\chi^2 \ge \frac{1}{1+\rho}\sum_{k=0}^{\infty} p_{2k+1}. \tag{A.3}$$

Since

$$\rho\sum_{k=0}^{\infty} p_{2k+1} \ge \sum_{k=0}^{\infty} p_{2k+2} = 1 - p_0 - \sum_{k=0}^{\infty} p_{2k+1},$$

inequality (A.3) becomes

$$\chi^2 \ge \frac{1}{1+\rho}\frac{1-p_0}{1+\rho},$$

and the first half of (A.2) is proved. The second half is proved in the same way from the inequality

$$\chi^2 \leq \frac{\rho}{1+\rho} \sum_{k=0}^{\infty} p_{2k}.$$

■

χ has the desirable property to be zero when X is an integer variable with a span strictly greater than 1, in which case the period of φ is less than 2π. Another desirable property would be that $\chi \to \infty$ when the moments of X are unbounded; since $\chi \leq 1/4$, this is obviously not possible here. That this "feature" is somehow unavoidable can be seen on the following example:

$$\begin{aligned}
\varphi(\theta) &\stackrel{\text{def}}{=} \frac{2+e^{i\theta}}{4} + \frac{1}{4} \sum_{k=2}^{\infty} \frac{e^{ik\theta}}{k(k-1)} \\
&= \frac{1+e^{i\theta}}{2} + (1-e^{i\theta}) \ln(1-e^{i\theta}).
\end{aligned}$$

The random variable having φ as characteristic function admits no finite moment of order greater or equal to 1, but no bound on $|\varphi|$ is substantially better than (A.1).

References

[1] A. BIRMAN AND Y. KOGAN, *Asymptotic evaluation of closed queueing networks with many stations*, Communications in Statistics—Stochastic Models, 8 (1992), pp. 543–563.

[2] W. FELLER, *An Introduction to Probability Theory and its Applications, Vol. II*, John Wiley & Sons, 2 ed., 1971.

[3] F. P. KELLY, *Reversibility and Stochastic Networks*, Wiley, 1979.

[4] C. KNESSL AND C. TIER, *Asymptotic expansions for large closed queueing networks*, Journal of the ACM, 37 (1990), pp. 144–174.

[5] Y. KOGAN, *Another approach to asymptotic expansions for large closed queueing networks*, Operations Research Letters, 11 (1992), pp. 317–321.

[6] Y. KOGAN AND A. BIRMAN, *Asymptotic analysis of closed queueing networks with bottlenecks*, in Proceedings of the International Conference on Performance of Distributed Systems and Integrated Communication Networks, T. Hasegawa, H. Takagi, and Y. Takahashi, eds., Kyoto, 1991, pp. 237–252.

[7] M. LOÈVE, *Probability Theory*, D. Van Nostrand Company, third ed., 1963.

[8] V. MALYSHEV AND A. YAKOVLEV, *Condensation in large closed Jackson networks*, Rapport de Recherche 1854, INRIA, Domaine de Voluceau, Rocquencourt B.P. 105 78153 Le Chesnay Cedex France, 1993.

[9] J. MCKENNA AND D. MITRA, *Integral representations and asymptotic expansions for closed Markovian queueing networks: Normal usage*, Bell System Technical Journal, 61 (1982), pp. 661–683.

[10] ——, *Asymptotic expansions and integral representations of moments of queue lengths in closed Markovian networks*, Journal of the ACM, 31 (1984), pp. 346–360.

[11] J. MCKENNA, D. MITRA, AND K. G. RAMAKRISHNAN, *A class of closed Markovian queueing networks: Integral representations, asymptotic expansions and generalizations*, tech. rep., AT&T Bell Laboratories, 1990.

[12] D. MITRA AND J. MCKENNA, *Asymptotic expansions for closed Markovian queueing networks with state dependent service rates*, Journal of the ACM, 33 (1986), pp. 568–592.

[13] R. F. SERFOZO, *Markovian network processes: Congestion-dependent routing and processing*, Queueing Systems, Theory and Applications, 5 (1989), pp. 5–36.

[14] P. WHITTLE, *Systems in Stochastic Equilibrium*, Wiley, 1986.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay 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 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)
Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399

RR_2513