

On Kelly Networks with Shuffling

T. Bonald[†], M-A. Tran^{*}

[†]France Telecom
thomas.bonald@orange-ftgroup.com

^{*}École Normale Supérieure
minh-anh.tran@polytechnique.org

January 21, 2008

Abstract

We consider Kelly networks with shuffling of customers within each queue. Specifically, each arrival, departure or movement of customer from one queue to another triggers a shuffle of the other customers at each queue. The shuffle distribution may depend on the network state and on the customer that triggers the shuffle. We prove that the stationary distribution of the network state remains the same as without shuffling. In particular, Kelly networks with shuffling have the product form. Moreover, the insensitivity property is preserved for symmetric queues.

Keywords: Product form, insensitivity, symmetric queues, shuffling.

1 Introduction

Since the pioneer work of Jackson [11], an intense research activity has been devoted to the study of queueing networks, with applications ranging from production lines to computer systems and communication networks [2, 8, 13]. Specific attention has been paid to product-form networks, whose stationary distribution can be evaluated explicitly, see e.g. [3, 4, 10, 17]. As originally proved by Kelly [12], the stationary distribution of such networks does not depend on the service time distribution beyond the mean for a large class of *symmetric* service disciplines, that includes the processor-sharing discipline and the preemptive-LIFO discipline considered by Baskett, Chandy, Muntz and Palacios [1]. This insensitivity property is of great practical interest since usual performance metrics like the mean sojourn time of customers can be evaluated without knowing the precise statistics of service requirements.

In the present paper, we consider Kelly networks with shuffling. When a customer arrives at or leaves a queue, the positions of the other customers present at this queue are permuted at random. Such random permutations have already been considered by Yashkov [18] and Daduna [6], see also Daduna, Schassberger [7, 5] and Yates [19, 20] for discrete-time queues. We here consider Kelly networks with arbitrary routes and assume that *any* movement of customer may trigger the shuffle of customers at *all* queues. The shuffle distribution may depend on the whole network state and on the customer that triggers the shuffle. Such a process does not enter the framework of Chao et. al. [4] where a customer or signal can only impact other queues through its propagation throughout the network. We prove that the stationary distribution of the network

state remains the same as without shuffling. In particular, Kelly networks with shuffling have the product form. Moreover, the insensitivity property is preserved for symmetric queues.

Following Kelly [12], we use the notion of customer *class* to represent arbitrary routes in the network. Without any loss of generality, we assume that each class is associated with a single queue. Customers of the same class require independent services of exponential distribution at this queue and become customers of another class or leave the network after service completion. The proof of the product form simply consists in verifying the associated partial balance equations. The general results of Miyazawa, Schassberger and Schmidt [14, 15, 16] showing the equivalence between partial balance and insensitivity for generalized semi-Markov schemes with reallocation may then apply to prove the insensitivity property of symmetric queues. We prefer a simple, direct proof based on the method of phases.

In the rest of the paper, we describe the considered queueing network in Section 2, prove the product form and the insensitivity property in Sections 3 and 4, respectively, and discuss some extensions in Section 5.

2 Network description

We consider an open network of J queues. In this section, we assume that all queues have unit service rates. The general case of state-dependent service rates is considered in Section 5. We denote by n_j the number of customers at queue j , by n the vector (n_1, \dots, n_J) and by e_j the unit vector with 1 in component j and 0 elsewhere. We refer to n as the *macro-state*. We define below the *micro-state* c that contains information about the classes and positions of customers in the queues.

Customer classes. We consider an arbitrary set of I customer classes. Without any loss of generality, we assume that each class is associated with a single queue. Denoting by C_j the set of classes associated with queue j , the sets C_1, \dots, C_J form a partition of $\{1, \dots, I\}$.

Class- i customers arrive according to a Poisson process of intensity ν_i . We refer to these arrivals as *external* arrivals. Class- i customers require independent, exponentially distributed services with mean $1/\mu_i$ at the associated queue. After service completion, a class- i customer becomes a class- k customer with probability p_{ik} , which corresponds to an *internal* arrival, and leaves the network with probability $p_i = 1 - \sum_{k=1}^I p_{ik}$. We assume that $\sum_{i=1}^I \nu_i > 0$ and that all customers eventually leave the network so that the arrival rate λ_i of class- i customers is the unique solution to the traffic equations:

$$\lambda_i = \nu_i + \sum_{k=1}^I \lambda_k p_{ki}. \quad (1)$$

The load due to class- i customers is then given by:

$$\rho_i = \frac{\lambda_i}{\mu_i}.$$

By increasing the number of classes, one may describe an arbitrary set of routes in the network, as in the original work of Kelly [12].

Service discipline. The service rate of queue j is shared by the n_j customers present at this queue as a function of their position in the queue, denoted by $l = 1, 2, \dots, n_j$. We denote by $\gamma_j(l, n)$ the fraction of the service rate directed to the customer in position l in macro-state n , with $\sum_{l=1}^{n_j} \gamma_j(l, n) = 1$; when this customer leaves the queue, customers previously in positions $l+1, l+2, \dots, n_j$ are shifted to positions $l, l+1, \dots, n_j-1$, respectively. A customer arriving at queue j in macro-state n moves into position l with probability $\delta_j(l, n + e_j)$, for all $l = 1, 2, \dots, n_j + 1$; customers previously in positions $l, l+1, \dots, n_j$ are shifted to positions $l+1, l+2, \dots, n_j+1$, respectively.

Queue j is said to be *symmetric* if for all macro-states n ,

$$\forall l = 1, \dots, n_j, \quad \delta_j(l, n) = \gamma_j(l, n). \quad (2)$$

If queue j is *not* symmetric, we assume that all customers in this queue have the same exponential service distribution, that is:

$$\forall i, k \in C_j, \quad \mu_i = \mu_k. \quad (3)$$

This is the case of FIFO queues for instance.

Micro-state. Let c_j be the sequence of n_j elements of C_j whose l -th element $c_j(l)$ is equal to the class of that customer in position l at queue j . We refer to the vector $c = (c_1, \dots, c_J)$ as the micro-state. We denote by 0 the micro-state corresponding to an empty network.

Let $\mathcal{C}(n)$ be the set of micro-states corresponding to the same macro-state n . For all $c \in \mathcal{C}(n)$, we denote by $c \oplus (i, l)$ the micro-state obtained from c by adding a class- i customer in position l to the associated queue j according to the shifting rule described above, for all $l = 1, 2, \dots, n_j + 1$. Similarly, we denote by $c \ominus (i, l)$ the micro-state obtained from c by deleting a class- i customer in position l from the associated queue j according to the above shifting rule, for all $l = 1, 2, \dots, n_j$ such that $c_j(l) = i$.

Shuffling. We refer to a *shuffle* in macro-state n as an element of $\mathcal{S}(n) = \mathcal{P}(n_1) \times \dots \times \mathcal{P}(n_J)$, where $\mathcal{P}(m)$ denotes the set of permutations of m elements, for all $m \geq 1$, and $\mathcal{P}(0)$ is the identity mapping on $\{\emptyset\}$. For any micro-state $c \in \mathcal{C}(n)$ and any shuffle $\sigma \in \mathcal{S}(n)$, we denote by $\sigma(c)$ the micro-state whose j -th component is equal to $\sigma_j(c_j)$, i.e., the positions of customers are permuted according to σ_j at queue j .

In the following, we assume that *any* movement of customer may trigger a shuffle of customers at each queue. The shuffle distribution may depend on the macro-state and on the customer that triggers the shuffle. For all macro-states n and classes i, k , let $\alpha_i(\cdot, n)$, $\beta_{ik}(\cdot, n)$, $\beta_i(\cdot, n)$ be arbitrary distributions on $\mathcal{S}(n)$. When a new class- i customer arrives in macro-state n , the other customers are permuted according to the shuffle distribution $\alpha_i(\cdot, n)$ immediately before the customer arrival. Specifically, the micro-state becomes $\sigma(c)$ with probability $\alpha_i(\sigma, n)$ immediately before the customer arrival. Similarly, when a class- i customer becomes a class- k customer after service completion in macro-state n , with $i \in C_j$, the other customers are permuted according to the shuffle distribution $\beta_{ik}(\cdot, n - e_j)$ immediately after the customer departure from queue j and before the arrival to the next queue. Finally, when a class- i customer leaves the network in macro-state n , with $i \in C_j$, the other customers are permuted according to the shuffle distribution $\beta_i(\cdot, n - e_j)$ after the customer departure from queue j .

3 Product form

The evolution of the micro-state defines a Markov process on a countable state space. The following result is proved in Appendix A:

Theorem 1 *The micro-state has the stationary measure:*

$$\pi(c) = \pi(0) \prod_{i=1}^I \rho_i^{x_i},$$

where $x_i = \sum_{l=1}^{n_j} 1_{\{c_j(l)=i\}}$ denotes the number of class- i customers at queue j , for all $i \in C_j$.

Note that the stationary measure of the micro-state is invariant by shuffling of customers at each queue. It only depends on the number of customers of each class. In particular, the macro-state has the stationary measure:

$$\bar{\pi}(n) = \sum_{c \in \mathcal{C}(n)} \pi(c) = \pi(0) \sum_{\substack{x_1, \dots, x_I: \\ \forall j, \sum_{i \in C_j} x_i = n_j}} \frac{n_j!}{\prod_{i \in C_j} x_i!} \prod_{i=1}^I \rho_i^{x_i} = \bar{\pi}(0) \prod_{j=1}^J \bar{\rho}_j^{n_j},$$

where $\bar{\rho}_j$ denotes the load of queue j :

$$\bar{\rho}_j = \sum_{i \in C_j} \rho_i.$$

We deduce that the network is stable if and only if $\bar{\rho}_j < 1$ for all $j = 1, \dots, J$, in which case the stationary distribution of the macro-state is given by:

$$\bar{\pi}(n) = \prod_{j=1}^J (1 - \bar{\rho}_j) \bar{\rho}_j^{n_j}.$$

The network has the product form.

4 Insensitivity property

Assume that the network is stable. In view of Theorem 1, the stationary distribution of the number of customers of each class at any given queue depends on the load of these classes only. Furthermore, the stationary distribution of the numbers of customers belonging to some subsets of classes forming a partition of the set of classes C_j at any given queue j depends on the total loads of these subsets of classes only. This shows the insensitivity of symmetric queues to *hyper-exponential* distributions of service requirements. To prove the insensitivity property of symmetric queues to arbitrary *phase-type* distributions of service requirements, that is mixtures of Erlang distributions, we replace the exponential distribution of service requirements of each class by an Erlang distribution with the same mean.

Specifically, assume that class- i customers require m_i independent exponential service phases with mean $1/(m_i \mu_i)$ at queue j , for all $i \in C_j$ and $j = 1, \dots, J$. The distribution of service requirements remains exponential for any non-symmetric queue j , so that $m_i = 1$ for all $i \in C_j$ in this case. Let s_j be the sequence of service stages of those customers present at queue j .

Specifically, if the customer in position l is of class i , that is $c_j(l) = i$, $s_j(l)$ is an element of $\{1, \dots, m_i\}$. We refer to the pair (c, s) , with $s = (s_1, \dots, s_J)$, as the *extended* micro-state. We still denote by 0 the extended micro-state corresponding to an empty network.

The evolution of the extended micro-state defines a Markov process on a countable state space. The following result is proved in Appendix B:

Theorem 2 *The extended micro-state has the stationary measure:*

$$\pi(c, s) = \pi(0) \prod_{i=1}^I \left(\frac{\rho_i}{m_i} \right)^{x_i},$$

where $x_i = \sum_{l=1}^{n_j} 1_{\{c_j(l)=i\}}$ denotes the number of class- i customers at queue j , for all $i \in C_j$.

Note that the stationary measure of the extended micro-state is independent of the service stage of each customer. It only depend on the number of customers of each class. Moreover, the stationary distribution of the micro-state remains unchanged. In particular, the stationary distribution of the numbers of customers belonging to some subsets of classes forming a partition of the set of classes C_j at any given queue j depends on the total loads of these subsets of classes only. This shows the insensitivity of symmetric queues to phase-type distributions. Such distributions are known to form a dense subset of the set of distributions with nonnegative support. The general insensitivity property then follows by continuity as in [4, 9].

5 Extensions

There are several possible extensions to these results. First, the same results apply to closed Kelly networks, as well as to mixed Kelly networks with both open and closed components. Second, the service rates may depend on the macro-state. Let $\phi_j(n)$ be the service rate of queue j in macro-state n . Assume that there is a positive function Φ on \mathbb{N}^J such that $\Phi(0) = 1$ and for all macro-states n such that $n_j \geq 1$,

$$\phi_j(n) = \frac{\Phi(n - e_j)}{\Phi(n)}.$$

This is the balance property of Kelly-Whittle networks [17]. The stationary measures of the micro- and macro-states become:

$$\pi(c) = \pi(0)\Phi(n) \prod_{i=1}^I \rho_i^{x_i} \quad \text{and} \quad \bar{\pi}(n) = \bar{\pi}(0)\Phi(n) \prod_{j=1}^J \bar{\rho}_j^{n_j}.$$

The insensitivity property is preserved for symmetric queues. Similarly, one may extend the result to state-dependent arrivals rates and routing probabilities, see e.g. [17].

As a final remark, we note that shuffles may additionally be triggered by an external point process at each queue (e.g. Poisson). It is indeed sufficient to consider additional queues whose role is to trigger shuffles (e.g. through customer arrivals at these queues). The intensity of shuffle events may depend on the network state (e.g. through state-dependent arrival rates).

Appendix

A Proof of Theorem 1

We prove that the stationary measure π of Theorem 1 satisfies the partial balance equations for all queues $j = 1, \dots, J$ and for the source of external arrivals.

Consider a micro-state $c \in \mathcal{C}(n)$ such that $c_j(l) = i$. The probability flux corresponding to the departure of the customer in position l from queue j in micro-state c is given by:

$$\pi(c)\mu_i\gamma_j(l, n). \quad (4)$$

Now the probability flux corresponding to the external arrival of a customer in position l at queue j leading to micro-state c is given by:

$$\sum_{\sigma \in \mathcal{S}(n-e_j)} \pi(c')\nu_i\alpha_i(\sigma, n-e_j)\delta_j(l, n),$$

where c' denotes the micro-state such that $c = c' \oplus (i, l)$. Using the fact that $\pi(c) = \pi(c')\rho_i$, we get the probability flux:

$$\pi(c)\frac{\nu_i}{\rho_i}\delta_j(l, n) \sum_{\sigma \in \mathcal{S}(n-e_j)} \alpha_i(\sigma, n-e_j) = \pi(c)\delta_j(l, n)\frac{\mu_i}{\lambda_i}\nu_i.$$

Similarly, the probability flux corresponding to the internal arrival of a customer in position l at queue j leading to micro-state c is given by:

$$\sum_{j'=1}^J \sum_{i' \in C_{j'}} \sum_{l'=1}^{n_{j'}+1_{\{j' \neq j\}}} \sum_{\sigma \in \mathcal{S}(n-e_j)} \pi(c')\mu_{i'}\gamma_{j'}(l', n+e_{j'}-e_j)p_{i'i}\beta_{i'i}(\sigma, n-e_j)\delta_j(l, n),$$

where c' denotes the micro-state such that $c = \sigma(c' \ominus (i', l')) \oplus (i, l)$. Using the fact that $\pi(c) = \pi(c')\rho_i/\rho_{i'}$, we get the probability flux:

$$\begin{aligned} \sum_{j'=1}^J \sum_{i' \in C_{j'}} \pi(c)\frac{\rho_{i'}}{\rho_i}\mu_{i'}p_{i'i}\delta_j(l, n) \sum_{l'=1}^{n_{j'}+1_{\{j' \neq j\}}} \gamma_{j'}(l', n+e_{j'}-e_j) \sum_{\sigma \in \mathcal{S}(n-e_j)} \beta_{i'i}(\sigma, n-e_j) \\ = \pi(c)\frac{\mu_i}{\lambda_i}\delta_j(l, n) \sum_{i'=1}^I \lambda_{i'}p_{i'i}. \end{aligned}$$

Summing both probability flux gives:

$$\pi(c)\frac{\mu_i}{\lambda_i}\delta_j(l, n) \left(\nu_i + \sum_{i'=1}^I \lambda_{i'}p_{i'i} \right).$$

In view of the traffic equations (1), the total probability flux corresponding to the arrival of a customer in position l at queue j leading to micro-state c is given by:

$$\pi(c)\mu_i\delta_j(l, n). \quad (5)$$

Now if queue j is symmetric, we deduce from (2) that the probability flux (4) and (5) are equal. If queue j is not symmetric, summing both probability flux over all positions $l = 1, \dots, n_j$ in the queue yields equality in view of (3).

To conclude the proof, we need to verify the partial balance equations for the source. Let $c \in \mathcal{C}(n)$. The probability flux corresponding to external arrivals in micro-state c is given by:

$$\pi(c) \sum_{i=1}^I \nu_i. \quad (6)$$

Now the probability flux corresponding to the departure of a customer from the network *leading* to micro-state c is given by:

$$\sum_{j'=1}^J \sum_{i' \in \mathcal{C}_{j'}} \sum_{l'=1}^{n_{j'}+1} \sum_{\sigma \in \mathcal{S}(n)} \pi(c') \mu_{i'} \gamma_{j'}(l', n + e_{j'}) p_{i'} \beta_{i'}(\sigma, n),$$

where c' denotes the micro-state such that $c = \sigma(c' \ominus (i', l'))$. Using the fact that $\pi(c) = \pi(c') / \rho_{i'}$, we get the probability flux:

$$\sum_{j'=1}^J \sum_{i' \in \mathcal{C}_{j'}} \pi(c) \rho_{i'} \mu_{i'} p_{i'} \sum_{l'=1}^{n_{j'}+1} \gamma_{j'}(l', n + e_{j'}) \sum_{\sigma \in \mathcal{S}(n)} \beta_{i'}(\sigma, n) = \pi(c) \sum_{i'=1}^I \lambda_{i'} p_{i'},$$

which is equal to the probability flux (6) in view of the traffic equations (1).

B Proof of Theorem 2

The proof is similar to that of Theorem 1. The only difference lies in the partial balance equations for symmetric queues.

Let $c \in \mathcal{C}(n)$ with $c_j(l) = i$ for some symmetric queue j . The probability flux corresponding to the completion of the service stage of the customer in position l at queue j in the extended micro-state (c, s) is given by:

$$\pi(c, s) \frac{\mu_i}{m_i} \gamma_j(l, n).$$

In particular, this probability flux does not depend on the service stage $s_j(l)$ of the customer. Now, by the same argument as in the proof of Theorem 1, we get the probability flux corresponding to the arrival of a customer in position l at queue j *leading* to micro-state (c, s) , with $s_j(l) = 1$:

$$\pi(c, s) \frac{\mu_i}{m_i} \delta_j(l, n).$$

Both probability flux are equal in view of (2).

References

- [1] F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios. Open, closed and mixed networks of queues with different classes of customers. *J. Assoc. Comput. Mach.*, 22:248–260, 1975.
- [2] D. Bertsekas and R. Gallager. *Data networks*. Prentice-Hall, 1987.

- [3] R. Boucherie and N. van Dijk. Product forms for queueing networks with state dependent multiple job transitions. *Adv. Appl. Probab.*, 23:152–187, 1991.
- [4] X. Chao, M. Miyazawa, and M. Pinedo. *Queueing Networks, Customers, Signals and Product Form solutions*. Wiley, Chichester, 1999.
- [5] H. Daduna. Stochastic networks with product form equilibrium. In D.N. Shanbhag and C.R. Rao, editors, *Stochastic Processes: Theory and Methods*, pages 309–364, Elsevier Science, Amsterdam, 2001.
- [6] H. Daduna. Queueing networks with discrete time scale: Explicit expressions for the steady state behavior of discrete time stochastic networks. *Lecture Notes in Computer Science*, 2046, Springer, Berlin, 2001.
- [7] H. Daduna and R. Schassberger. Networks of queues in discrete time. *Zeitschrift fuer Operations Research ZOR*, 27:159–175, 1983.
- [8] M.C. Fu and M.K. Govil. Queueing theory in manufacturing: A review. *Journal of Manufacturing Systems*, 18:214–240, 1999.
- [9] A. Hordijk and S. Schassberger. Weak convergence of generalized semi-markov processes. *Stochastic Process. Appl.*, 12:271–291, 1982.
- [10] A. Hordijk and N. van Dijk. Adjoint processes, job local balance and insensitivity for stochastic networks. *Bull: 44 Session Int. Stat. Inst.*, 50:776–788, 1982.
- [11] J.R. Jackson. Networks of waiting lines. *Operations Research*, 5:518–521, 1957.
- [12] F.P. Kelly. *Reversibility and Stochastic Networks*. Wiley, New York, 1979.
- [13] L. Kleinrock. *Queueing Systems*, volume 2. Wiley, 1975.
- [14] M. Miyazawa. Insensitivity and product-form decomposability of reallocatable GSMP. *Adv. Appl. Probab.*, 25:415–437, 1993.
- [15] M. Miyazawa, S. Schassberger, and S. Schmidt. On the structure of an insensitive generalized semi-Markov process with reallocation and point-process input. *Adv. Appl. Probab.*, 27:203–225, 1995.
- [16] R. Schassberger. The insensitivity of stationary distributions in networks of queues. *Adv. Appl. Probab.*, 10:906–912, 1978.
- [17] R.F. Serfozo. *Introduction to Stochastic Networks*. Springer, Berlin, 1999.
- [18] S.F. Yashkov. Properties of invariance of probabilistic models of adaptive scheduling in shared use systems. *Automatic control and computer science*, 14:46–51, 1980.
- [19] R.D. Yates. *High speed round-robin queueing networks*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1990.
- [20] R.D. Yates. Analysis of discrete time queues via the reversed process. *Queueing Systems and Their Applications*, 18:107–116, 1994.