Mathematical Methods in Operations Research 2005
Special issue in honor of *Arie Hordijk*

On Product Form Tandem Structures

Nico M. Van Dijk

University of Amsterdam, Department of Economic Sciences and Econometrics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Received: January 2005 / Revised version: March 2005

Abstract

This paper is written in honour to A. Hordijk.

It establishes product form results for a generic and instructive multi-class tandem queue with blocking, to which A. Hordijk has directly and indirectly contributed.

First, a sufficient and necessary product form characterization is provided. Next, three special cases are briefly presented. These illustrate the possibility of product forms despite finite capacity constraints (blocking), unproportional processor sharing mechanisms and resource contentions (such as for access control).

The results are partially new and of interest for present-day applications. In essence these rely upon the pioneering work by A. Hordijk.

Keywords. Tandem Queues * Product Forms

AMS Subject Classification. 60K25, 90B22

1 Introduction

Tandem structures are most naturally arising in a variety of practical queueing applications ranging from classical production line structures, circuit switch networks up to present-day applications as call centers and internet.

This paper therefore will investigate the existence of so-called product form solutions for an instructive yet generic example of (two) interdependent service stations (a tandem queue) and (two) interdependent job-classes. First, a general sufficient and necessary condition will be derived to conclude a product form solution. Next, three special cases are dealt with which show the possibility of product forms also in the presence of:

– Intermediate blocking and delays
– A load dependent service sharing over the stations
– Job-class interdependent access (and departure) blocking

In addition, the insensitivity phenomenon is addressed briefly. The results are still highly actual and in the combined form as presented, which can be regarded as a multi-class extension and combination of earlier results that Arie Hordijk has contributed to, have not yet been reported.

2 General Product from Characterization

Purely for instructional and illustrative purpose, while still preserving the generic structure of consecutive service stages and job-class dependent service interactions, we restrict the presentation to a tandem structure with two service stations, indexed by $i = 1, 2$, and two job-classes indexed, by $r = 1, 2$. Let the vector $\boldsymbol{n}$ denote the state of the system as specified by:

$$\boldsymbol{n} = (\boldsymbol{n}_1, \boldsymbol{n}_2) \quad \text{where}$$
$$\boldsymbol{n}_i = (n_i^1, n_i^2) \quad \text{with } n_i^r \text{ the numbers of class-}r\text{ jobs at station } i.$$

Parametrization. Under the assumption of exponential interarrival and service times, the following (exponential) parameters are then involved:

λ^r : arrival rate of class-r jobs at station 1.
μ_i^r : service parameter (of the exponential service requirement) of a class-r job at station i.
$\boldsymbol{f}_i^r(\boldsymbol{n}_1, \boldsymbol{n}_2)$: total service capacity provided to class-r jobs at station i when the system is in state $\boldsymbol{n} = (\boldsymbol{n}_1, \boldsymbol{n}_2)$.
$\boldsymbol{b}_{ij}^r(\boldsymbol{n}_1, \boldsymbol{n}_2)$: probability that a transition of a class-r job from station i to station $j = i + 1$ is accepted to change the system state from $\boldsymbol{n} + e_i^r$ into $\boldsymbol{n} + e_j^r$ (see notation below) where $i = 0$ indicates an arrival at the system and $j = 0$ a departure from the system.

Notation. As we need to keep track of the total numbers of jobs also let,

$$\boldsymbol{t} = (n_1, n_2) \quad \text{with } n_i = n_i^1 + n_i^2, \quad i = 1, 2;$$
$$\boldsymbol{z} = (z^1, z^2) \quad \text{with } z^r = n_1^r + n_2^r, \quad r = 1, 2.$$

Furthermore, we use the unit vector e_i^r, e_i, e^r with a + or a − sign for the corresponding states to denote one job more or less. Hence, $\boldsymbol{n} + e_i^r$ for the state with one class-r job more of at station i and $\boldsymbol{n}$ - e_j^r for the state with one class-r job less at station j, and similarly $\boldsymbol{z} + e^r$ or $\boldsymbol{z} - e^r$. We also use the convention that $\boldsymbol{n} + e_0^r = \boldsymbol{n} - e_0^r = \boldsymbol{n}$.

Steady state distribution

Under natural ergodicity assumptions for its existence, let $\pi(\boldsymbol{n}_1, \boldsymbol{n}_2)$ denote the corresponding steady state distribution. We aim to investigate and characterize the analytical feasibility of this distribution. To this end, it suffices to find the (unique probability) solution of the global balance (or steady state Kolmogorov) equations. With $\pi(\boldsymbol{n}_1, \boldsymbol{n}_2) = 0$ for all $(\boldsymbol{n}_1, \boldsymbol{n}_2) \notin C$ where C is the set of admissible states $(\boldsymbol{n}_1, \boldsymbol{n}_2)$, these are given by:

$$\left\{\begin{array}{l} \pi(\boldsymbol{n})\mu_1^1 \boldsymbol{f}_1^1(\boldsymbol{n})\, \boldsymbol{b}_{12}^1(\boldsymbol{n}-e_1^1)+ \\ \pi(\boldsymbol{n})\mu_1^2 \boldsymbol{f}_1^2(\boldsymbol{n})\, \boldsymbol{b}_{12}^2(\boldsymbol{n}-e_1^2)+ \\ \pi(\boldsymbol{n})\mu_2^1 \boldsymbol{f}_2^1(\boldsymbol{n})\, \boldsymbol{b}_{20}^1(\boldsymbol{n}-e_2^1)+ \\ \\ \pi(\boldsymbol{n})\mu_2^2 \boldsymbol{f}_2^2(\boldsymbol{n})\, \boldsymbol{b}_{20}^2(\boldsymbol{n}-e_2^2)+ \\ \\ \pi(\boldsymbol{n})\lambda^1 \boldsymbol{b}_{01}^1(\boldsymbol{n})+ \\ \pi(\boldsymbol{n})\lambda^2 \boldsymbol{b}_{01}^2(\boldsymbol{n}) \end{array}\right\} = \left\{\begin{array}{l} \pi(\boldsymbol{n}-e_1^1)\lambda^1 \boldsymbol{b}_{01}^1(\boldsymbol{n}-e_1^1)+ \\ \pi(\boldsymbol{n}-e_1^2)\lambda^2 \boldsymbol{b}_{01}^2(\boldsymbol{n}-e_1^2)+ \\ \pi(\boldsymbol{n}+e_1^1-e_2^1)\mu_1^1 \boldsymbol{f}_1^1(\boldsymbol{n}+e_1^1-e_2^1) \\ \qquad\qquad \boldsymbol{b}_{12}^1(\boldsymbol{n}-e_2^1)+ \\ \pi(\boldsymbol{n}+e_1^2-e_2^2)\mu_1^2 \boldsymbol{f}_1^2(\boldsymbol{n}+e_1^2-e_2^2) \\ \qquad\qquad \boldsymbol{b}_{12}^2(\boldsymbol{n}-e_2^2)+ \\ \pi(\boldsymbol{n}+e_2^1)\mu_2^1 \boldsymbol{f}_2^1(\boldsymbol{n}+e_2^1)\, \boldsymbol{b}_{20}^1(\boldsymbol{n})+ \\ \pi(\boldsymbol{n}+e_2^2)\mu_2^2 \boldsymbol{f}_2^2(\boldsymbol{n}+e_2^2)\, \boldsymbol{b}_{20}^2(\boldsymbol{n}) \end{array}\right\}$$

We cannot expect an analytic solution unless we might be able to decompose these global equations into a more detailed balance equation for each i and r separately. More precisely, with $\mu_0^r = \lambda^r$, $f_0^r(\boldsymbol{n}) \equiv 1$, and $i-1=0$ for $i=1$ and $i+1=0$ for $i=2$, for each i and r that is:

$$\left\{\begin{array}{l} \pi(\boldsymbol{n})\mu_i^r \boldsymbol{f}_i^r(\boldsymbol{n})\, \boldsymbol{b}_{i(i+1)}^r(\boldsymbol{n}-e_i^r) = \\ \pi(\boldsymbol{n}-e_i^r+e_{i-1}^r)\mu_{i-1}^r \boldsymbol{f}_{i-1}^r(\boldsymbol{n}-e_i^r+e_{i-1}^r)\, \boldsymbol{b}_{(i-1)i}^r(\boldsymbol{n}-e_i^r) \end{array}\right\} \tag{1}$$

Adjoint Markov chain

To investigate the existence of a solution for (1), consider a continuous-time Markov chain at C, which will be called an *adjoint Markov chain*, with transition rates $\bar{q}(\boldsymbol{n}+e_i^r, \boldsymbol{n}+e_j^r)$ for a change from a state $\boldsymbol{n}+e_i^r$ into $\boldsymbol{n}+e_j^r$ given by:

$$\bar{q}(\boldsymbol{n}+e_i^r, \boldsymbol{n}+e_j^r) = \begin{cases} \boldsymbol{f}_i^r(\boldsymbol{n}+e_i^r)\ \boldsymbol{b}_{i(i+1)}^r(\boldsymbol{n}), & j=i+1 \\ \boldsymbol{f}_i^r(\boldsymbol{n}+e_i^r)\ \boldsymbol{b}_{(i-1)i}^r(\boldsymbol{n}), & j=i-1 \end{cases} \tag{2}$$

(Note that this chain coincides with the parametrization of the original tandem system in the natural flow direction $i \to i+1$. In contrast though, also a flow in opposite direction is constructed).

Let $H(\boldsymbol{n})$ denote the corresponding steady state distribution at C. This Markov chain is called reversible if for any pair of states $\boldsymbol{n}, \boldsymbol{n}'$:

$$H(\boldsymbol{n})\ \bar{q}(\boldsymbol{n}, \boldsymbol{n}') = H(\boldsymbol{n}')\ \bar{q}(\boldsymbol{n}', \boldsymbol{n}) \tag{3}$$

Result 1. *There exists a solution $\pi(\boldsymbol{n})$ which satisfies the station-class balance relation (1) for each (i, r) if and only if the adjoint Markov chain is reversible. In that case, with c a normalizing constant:*

$$\pi(\boldsymbol{n}) = c\,H(\boldsymbol{n}) \prod_i \prod_r \left[\frac{1}{\mu_i^r}\right]^{n_i^r} \tag{4}$$

Result 1 is verified directly by substitution in the balance equations (1) for each (i, r) separately. It characterizes the existence of a product form solution by means of *reversibility*, despite the fact that the systems itself is *non-reversible.*

Reversibility characterization. The major advantage of result 1 is that it enables one to verify the existence of a (product form) solution of the form (4), by simply investigating the existence of a reversible solution $H(\cdot)$. This in turn, can be verified by the so-called Kolmogorov criterion for reversibility (see [8]). In the present case that is, as based upon just the transition rates (2). This can also be achieved by checking whether for any $\boldsymbol{n} \in C$:

$$H(\boldsymbol{n}) = c \prod_{k=0}^{K-1} \left[\frac{\bar{q}(\boldsymbol{n}_k \to \boldsymbol{n}_{k+1})}{\bar{q}(\boldsymbol{n}_{k+1} \to \boldsymbol{n}_k)}\right] \tag{5}$$

for any path $\boldsymbol{n}_0 \to \boldsymbol{n}_1 \to \ldots \to \boldsymbol{n}_K \to \boldsymbol{n}$ (for which the denominator is positive). Either of these checks in turn can generally be reduced to basic cycles or short paths that directly suggest a necessary form of $H(\cdot)$ and a decomposition in a service and routing component, satisfying:

$$\left[\frac{H(\boldsymbol{n} + e_i^r)}{H(\boldsymbol{n} + e_j^r)}\right] = \left[\frac{\boldsymbol{f}_j^r(\boldsymbol{n} + e_j^r)}{\boldsymbol{f}_i^r(\boldsymbol{n} + e_i^r)}\right] \left[\frac{\boldsymbol{b}_{ji}^r(\boldsymbol{n})}{\boldsymbol{b}_{ij}^r(\boldsymbol{n})}\right] \tag{6}$$

Literature

The concept of an adjoint (artificial) Markov chain to characterize the existence (or not) of a product form solution has first been introduced and exploited in Hordijk and Van Dijk [1]. A further extension and generalization to job-locally balanced networks has been developed in Hordijk and van Dijk [3]. And also extensions to partial balance results such as for decomposition and multiple movements have been established (e.g. [11], [12]).

The product form characterization presented above can be regarded as a multi-class extension of Hordijk and Van Dijk [1] and Van Dijk [10]. This multi-class extension may still lead to novel applications of present-day practical interest. In essence though, these results rely upon the pioneering work in the aforementioned references by *Arie Hordijk*.

Remark (Insensitivity) A PF related aspect to which A. Hordijk also contributed is that of *insensitivity* (Hordijk and Schassberger [7], Hordijk and Van Dijk [3], [4]). Most notably, the class of symmetric and insensitive disciplines (cf. [8], [13]) was extended in Hordijk and Van Dijk [2].

3 Special Cases

In this section some special cases will be presented for each of which the reversibility condition of result 1 is satisfied so that the product form applies. The verification of this condition with the specific form of $H(\cdot)$ presented, is referred to or left to the reader. In each case, we focus on only one specific aspect but any combination of them is easily concluded. The functions not mentioned are standardly parameterized by:

$$\begin{cases} \boldsymbol{b}_{ij}^r(\boldsymbol{n}) = 1 \qquad \forall\, i,\, j,\, r,\, \boldsymbol{n} \quad \text{and} \\ \boldsymbol{f}_i^r(\boldsymbol{n}_1, \boldsymbol{n}_2) = f_i^r(n_i^r) \qquad \forall\, i,\, r. \end{cases} \tag{7}$$

Let

$$F(\boldsymbol{n}) = \prod_i \prod_r \left[\prod_{k=1}^{n_i^r} f_i^r(k) \right]^{-1} \tag{8}$$

Case 1 (Blocking by total station populations) Assume that

$$\boldsymbol{b}_{i(i+1)}^r(\boldsymbol{n}_1, \boldsymbol{n}_2) = \boldsymbol{b}_{i+1}(n_1, n_2)$$

As can be concluded directly from the proof in Hordijk and Van Dijk [1], the reversibility condition (3) now reduces to:

$$\begin{aligned} &\boldsymbol{b}_1(n_1+1, n_2)\boldsymbol{b}_2(n_1, n_2+1)\boldsymbol{b}_0(n_1, n_2) \\ &= \boldsymbol{b}_1(n_1, n_2+1)\boldsymbol{b}_2(n_1, n_2)\boldsymbol{b}_0(n_1+1, n_2) \end{aligned} \tag{9}$$

For the natural situation with finite capacity constraints N_i condition (9) is clearly violated by a standard blocking function: $\boldsymbol{b}_i(n_1, n_2) = \mathbf{1}_{(n_i+1 \le N_i)}$. However, with these finite capacity constraints, condition (9) can still be satisfied by setting $\boldsymbol{b}_i(n_1, n_2) = \mathbf{1}_{(n_i \le N_i,\ n_{i+1} \le N_{i+1})}$ (where $N_0 = \infty$) with

$$\begin{gathered} H(\boldsymbol{n}) = \mathbf{1}_{(\boldsymbol{n} \in C)} F(\boldsymbol{n}) \quad \text{with} \\ C = \{\boldsymbol{n} | n_1 \le N_1, n_2 \le N_2, n_1 + n_2 \ne N_1 + N_2\}. \end{gathered} \tag{10}$$

This type of product form result, as initiated in Hordijk and Van Dijk [1], has also led to product form bounds for non-product form systems (e.g. [5], [6] for overflow and [10], [15], [16], [17] for production line structures).

Remark (Reversible routing) Note that a PF condition of a reversible routing as in [8] and [14] is necessarily violated by any tandem structure.

Case 2 (Unproportional Processor Sharing) As an extension of standard processor sharing disciplines for one service location, in present-day service structures, such as internet (cf. [19]), a single service entity may have to share its capacity over multiple service stations or job-classes as parameterized by:

$$\boldsymbol{f}_i^r(\boldsymbol{n}_1, \boldsymbol{n}_2) = \Phi(n_1 + n_2)\, \boldsymbol{s}_i(n_i|n_1 + n_2), \qquad i = 1, 2, \quad r = 1, 2 \tag{11}$$

where $\Phi(\cdot)$ represents the total service capacity of the service entity and where $\boldsymbol{s}_i(\cdot|\cdot)$ represents the fraction of this capacity allocated to station i. A processor sharing which would allocate capacity over *both* stations proportional to the workloads present as recently reported in ([18]) is hereby included by: $\boldsymbol{s}_i(n_i|n_1+n_2) = n_i/(n_1+n_2)$. But also *unproportional* sharing functions over both stations might still retain the invariance condition (5), for example:

$$\boldsymbol{s}_i(n_i|n_1 + n_2) = \begin{cases} \frac{2}{3} \quad i = 1, & \frac{1}{3} \quad i = 2, \quad n_1 > n_2 \\ \frac{1}{3} \quad i = 1, & \frac{2}{3} \quad i = 2, \quad n_1 < n_2 \\ \frac{1}{3} \quad i = 1, 2 & \qquad\qquad n_1 = n_2 \end{cases} \tag{12}$$

(Note that a capacity of $\frac{1}{3}$ is lost when $n_1 = n_2$). Condition (5) or (6) can now be verified with

$$H(\boldsymbol{n}) = \left[\prod_{k=1}^{n_1+n_2} \Phi(k)\right]^{-1} \left[2^{\max(n_1,n_2)}\right]^{-1} [3]^{n_1+n_2} \tag{13}$$

This unproportional processor sharing product form has not been reported (cf. [18]) and is still topical such as for internet modeling (cf. [19]).

Similar unreported product form results (e.g. as of the form in (12)) can also be concluded when the processor capacity is shared over different job-classes, such as for voice and data in circuit switched or mobile communications.

Case 3 (Access blocking) As another present day feature in (tele)communications different access or allocation schemes and limitations might be in order for different job-classes, such as illustrated below for circuit switched (direct end-to-end connections) or packet switched (via intermediate stages) communication structures. To this end, with z^r the total number of class-r jobs, let

$$\begin{cases} \boldsymbol{b}_{0\,1}^r(\boldsymbol{n}) = \boldsymbol{A}^r(z^1, z^2) \\ \boldsymbol{b}_{1\,0}^r(\boldsymbol{n}) = \boldsymbol{b}_{2\,0}^r(\boldsymbol{n}) = \boldsymbol{D}^r(z^1, z^2) \end{cases} \tag{14}$$

Here
$\boldsymbol{A}^r(\cdot,\cdot)$ is to be seen as an access (or arrival) blocking function while
$\boldsymbol{D}^r(\cdot,\cdot)$ as a delay function for class-r jobs. (One may standardly think of
$\boldsymbol{D}^r(\cdot,\cdot) \equiv 1$ but below also an example will be given with $\boldsymbol{D}^r(\cdot,\cdot) = 0$).

Let $v(z^1, z^2) = (1, \ldots 1, 2, \ldots, 2)$ denote the vector with the first z^1 components equal to 1 and the next z^2 equal to 2. Furthermore, for a vector $(r_1, r_2, \ldots, r_t)$ with $r_i = 1$ or 2 for all i, let $z^r(r_1, r_2, \ldots, r_t)$ denote the number of components equal to r $(r = 1, 2)$.

The invariance condition (5) can then be shown to be equivalent to the existence of a function $P(z^1, z^2)$ such that for any configuration (z^1, z^2):

$$P(z^1, z^2) = \prod_{k=1}^{z^1+z^2} \left[\frac{\boldsymbol{A}^{r_k}\left(z^1(r_1, r_2, ..., r_t), z^2(r_1, r_2, ..., r_t)\right)}{\boldsymbol{D}^{r_k}\left(z^1(r_1, r_2, ..., r_t), z^2(r_1, r_2, ..., r_t)\right)} \right] \tag{15}$$

for any permutation $(r_1, r_2, \ldots, r_{z^1+z^2})$ from $v(z^1, z^2)$ for which the denominator is positive and under the assumption that there is at least one permutation for which the denominator is positive. Furthermore, in this case $H(\boldsymbol{n}) = F(\boldsymbol{n})P(z^1, z^2)$.

Example 1 : A rich class of (tele)communication examples (e.g. circuit switch), in which case (15) is trivially satisfied with $P(\cdot, \cdot) \equiv 1$ is provided by:

$$\left\{ \begin{array}{c} \boldsymbol{A}^r(z^1, z^2) = \mathbf{1}_{(\boldsymbol{z}+e^r \in C)} \text{ and } \boldsymbol{D}^r(\boldsymbol{z}) \equiv 1 \text{ provided:} \\ \boldsymbol{z} \in C \Rightarrow \boldsymbol{z} - e^r \in C \ (\forall r) \ \text{ i.e.} \\ C \text{ is coordinate convex} \end{array} \right\} \tag{16}$$

Example 2 : As some sort of priority example for which (15) is also satisfied with $P(\cdot, \cdot) \equiv 1$, type 2 jobs might not be accepted and handled (block and stop type 2 services) where the workload of type 1 jobs exceeds some threshold M^1 by:

$$\left\{ \begin{array}{c} \boldsymbol{A}^2(z^1, z^2) = \boldsymbol{D}^2(z^1, z^2) = \mathbf{1}_{(z^1 \leq M^1)} \\ H(\boldsymbol{n}) = F(\boldsymbol{n}) \end{array} \right\} \tag{17}$$

Retrospection

The pioneering work of Arie Hordijk in the field of stochastic and queueing networks has led to a number of research directions and practical results:

- *partial balance insights for product form characterizations,*
- *secure analytic performance bounds and*
- *insensitivity results.*

His intuitive and fundamental probabilistic approach has been essential to establish these results. The author is most grateful to him for the stimulating collaboration in these research directions. Some special locations and occasions at which crucial steps for this research have been made are deeply kept in memory.

References

[1] Hordijk A and Van Dijk NM (1981) Networks of queues with blocking. In: Kylstra KJ (ed.) Performance 81. North-Holland, Amsterdam, pp. 51–65
[2] Hordijk A and Van Dijk NM (1982) Stationary probabilities for networks of queues. In: Disney L and Ott TJ (eds.) Applied Probability-Computer Science: The Interface, Vol. II. Birkhäuser, Basel, pp. 423–451
[3] Hordijk A and Van Dijk NM (1983) Networks of queues, Part I: Job-local-balance and the adjoint process, Part II: General routing and service characteristics. In: Lecture Notes in Control and Information Sciences, Vol. 60. Springer-Verlag, Berlin, pp. 158–205
[4] Hordijk A and Van Dijk NM (1983) Adjoint processes, job-local-balance and insensitivity for stochastic networks. 44th Session Int. Statist. Inst. 50: 776–788
[5] Hordijk A and Ridder A (1987) Stochastic inequalities for an overflow model. J. Appl. Probab. 24: 696–708
[6] Hordijk A and Ridder A (1988) Approximating sensitive queueing networks by reversible Markov chains. Computer Performance and Reliability, 105–117
[7] Hordijk A and Schassberger R (1982) Weak convergence of generalized semi-Markov processes. Stochastic Process. Appl. 12: 271–291
[8] Kelly FP (1979) Reversibility and Stochastic Networks. Wiley, New York
[9] Kelly FP (1991) Loss Networks. Ann. Appl. Probab. 1: 319–378
[10] Van Dijk NM (1993) Queueing networks and product forms. Wiley
[11] Boucherie RJ and Van Dijk NM (1991) Product forms for queueing networks with multiple job transitions. Adv. in Appl. Probab. 23: 152–187
[12] Boucherie RJ and Van Dijk NM (1993) A generalization of Norton's theorem. Queueing Systems 13: 251–289
[13] Chandy KM and Martin AJ (1983) A characterization of product-form queueing networks. J. Assoc. Comput. Mach. 30: 286–299
[14] Pittel B (1979) Closed exponential networks of queues with saturation. Math. Oper. Res. 4: 357–378
[15] Van Dijk NM (1998) Bounds and error bounds for queueing networks. Ann. Oper. Res. 79: 295–319
[16] Van Dijk NM and Miyazawa M (1997) Error bounds on a practical approximation for finite tandem queues. Oper. Res. Lett. 21: 201–208
[17] Van Dijk NM and Van der Sluis E (2004) Simple product-form bounds for queueing networks with finite clusters. Ann. Oper. Res. 113: 175–195
[18] Bonald T and Proutiére A (2002) Insensitivity in PS-networks. Perf. Eval. 49: 193–205
[19] Weij W van der (2005) Master Thesis, University of Amsterdam