A successive censoring algorithm for a system of connected QBD-processes

Niek Baer[1], Ahmad Al Hanbali[2], Richard J. Boucherie[1], Jan-Kees van Ommeren[1]

[1]Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands

[2]School of Management and Governance, Department of Industrial Engineering and Business Information Systems, University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands

{n.baer, a.alhanbali, r.j.boucherie, j.c.w.vanommeren}@utwente.nl

December 30, 2013

Abstract

We consider a Markov Chain in which the state space is partitioned into sets where both transitions within sets and between sets have a special structure. Transitions *within* each set constitute a finite Quasi-Birth-and-Death-process, and transitions *between* sets are restricted to four types of transitions. We present a successive censoring algorithm, based on Matrix Analytic Methods, to obtain the stationary distribution of this system of connected QBD-processes.

Keywords: successive censoring algorithm, Matrix Analytic Methods, connected QBD-processes, steady state analysis, exact aggregation/disaggregation

1 Introduction

We consider a class of Markov Chains in which the state space can be partitioned into sets. Transitions within each set constitute a finite Quasi-Birth-and-Death process (QBD), and the transitions between sets follow a special structure. This way, we create a system of connected QBD-processes on the whole state space. Such a system of connected QBD-processes often occurs in queueing systems with hysteresis in both traffic [1] and telecommunication systems [15]. To obtain the stationary distribution of such a Markov Chain, we present a successive censoring algorithm based on the censoring algorithm by Kemeny and Snell [11] for discrete time Markov Chains. This successive censoring algorithm allows for easy computation of the stationary distribution of a network of multi-threshold queues [1]. Until now, this was only possible by directly solving $\pi \boldsymbol{Q} = \boldsymbol{0}$.

The concept of the successive censoring algorithm is not new, Gaver, Jacobs and Latouche [6], use the same approach to determine the stationary distribution of a Level Dependent Quasi-Birth-and-Death process (LDQBD) with a finite number of levels. Our work extends the work of Gaver, Jacobs and Latouche [6] from censoring one level per iteration to censoring a complete QBD-process per iteration. The censoring algorithm [11] also forms the base for the folding algorithm in Ye and Li [19] and Li and Sheng [16], where the stationary distribution of a finite QBD was obtained by sequentially splitting (and renumbering) the state space in odd and even numbered sets, followed by application of the censoring algorithm to the two resulting subsets.

In the literature, the censoring algorithm is also called exact aggregation/disaggregation algorithm in which the state space is aggregated to obtain a smaller (and easier to solve) Markov Chain. The stationary distribution for this aggregated Markov Chain is then disaggregated to obtain the stationary distribution of the full Markov Chain. Similar exact aggregation/disaggregation algorithms can be found in literature. Most recent is the work of Katehakis and Smit [9] and Katehakis, Smit and Spieksma [10]. In [9], a Markov Chain is studied in which the state space is partitioned in sets, without any restrictions on the transitions within a set. In their successive lumping procedure it is crucial that a set contains a single entrance state, a state through which the set can be reached from other sets. Our work extends this aggregation method by allowing multiple entrance states, under restriction that the transitions within a set form a QBD. The work in [9] is applied to Quasi-Skip Free Processes to the left in [10] where it is assumed that lower levels are entered via one entrance state only. The single entrance states in [9, 10] are called *mandatory* states in Kim and Smith [12] and *input* states in Feinberg and Chui [5] in which a *parallel* lumping procedure was introduced.

For a thorough overview and comparison of several aggregation/disaggregation algorithms see Cao and Stewart [3], Haviv [7], Kafeety, Meyer and Stewart [8] and Rogers and Plante [18].

Section 2 introduces the system of connected QBD-processes and specifies the exact restrictions on the transitions between the QBD-processes. In Section 3 we present the successive censoring algorithm to determine the stationary distribution of the system of connected QBD-processes. In Section 4 we give an algorithm which determine if the successive censoring algorithm can be applied for a given Markov Chain. It also prepares the Markov Chain such that the successive censoring algorithm can be used directly. A complete overview and a demonstration of the successive censoring algorithm is given in Section 5. We also determine the complexity of the successive censoring algorithm in Section 5. Section 6 gives concluding remarks.

2 Model Description

Consider an irreducible and positive recurrent continuous time Markov Chain $\mathcal{X}$ with finite state space $\mathcal{S}$ that can be partitioned into sets ω_k, $k = 1, \ldots, S$. Each set ω_k contains a L_s levels labelled $l_1, \ldots, l_{L_k}$, each with equal number of P_s phases labelled $p_1, \ldots, p_{P_k}$. A state is denoted by the three-tuple (s, l, p) describing the *set*, *level* and *phase*. Let $\boldsymbol{Q}$ be its infinitesimal generator in which the states are ordered lexicographically:

- $(1, 1, 1), \ldots, (1, 1, P_1), \ldots, (1, L_1, 1), \ldots, (1, L_1, P_1)$
- $\cdots$
- $(S, 1, 1), \ldots, (S, 1, P_S), \ldots, (S, L_S, 1), \ldots, (S, L_S, P_S)$

The transitions *within* ω_i constitute a Quasi-Birth-and-Death process (QBD), making $\boldsymbol{Q}_{i,i}$ a tri-diagonal block matrix:

$$\boldsymbol{Q}_{i,i} = \begin{bmatrix} \boldsymbol{L}_1^i & \boldsymbol{F}^i & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{B}^i & \boldsymbol{L}^i & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \boldsymbol{L}^i & \boldsymbol{F}^i \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{B}^i & \boldsymbol{L}_{L_i}^i \end{bmatrix}. \tag{1}$$

Here $\boldsymbol{F}^i$ describes the transitions from l_a to l_{a+1}, $\boldsymbol{B}^i$ describes the transitions from l_{a+1} to l_a, and $\boldsymbol{L}^i$ describes the transitions within l_a for $a \neq 1, l_{L_i}$. The submatrices $\boldsymbol{L}_1^i$ and $\boldsymbol{L}_{L_i}^i$, which can differ from $\boldsymbol{L}^i$, describe the boundary transitions within l_1 and l_{L_i}.

The transitions *between* two sets ω_j and ω_k, $j \neq k$, are governed by two sets of conditions, labelled *direct* and *indirect* conditions. These conditions ensure that the QBD-structure of each set is maintained throughout the successive censoring algorithm that will be introduced in Section 3.

We denote by $\boldsymbol{Q}_{j,k}$, $j, k = 1, \ldots, S$, the submatrix of $\boldsymbol{Q}$ with transitions from ω_j to ω_k, and by $\left[\boldsymbol{Q}_{j,k}\right]_{a,b}$, $a = 1 \ldots, L_j$ and $b = 1 \ldots, L_k$, the submatrix of $\boldsymbol{Q}_{j,k}$ with transitions from l_a in ω_j to l_b in ω_k.

Definition 1. The direct conditions describe the one step transitions between ω_j and ω_k. We define five sets of transitions that can occur between ω_j and ω_k for $(j < k)$:

$T1$**:** Transitions from any level l in ω_j to only the *first* level l_1 in ω_k and back, i.e., $\left[\boldsymbol{Q}_{j,k}\right]_{a,b} = \boldsymbol{0}$ and $\left[\boldsymbol{Q}_{k,j}\right]_{b,a} = \boldsymbol{0}$ if $b \neq 1$.

$T2$**:** Transitions from any level l in ω_j to only the *last* level l_{L_k} in ω_k and back, i.e., $\left[\boldsymbol{Q}_{j,k}\right]_{a,b} = \boldsymbol{0}$ and $\left[\boldsymbol{Q}_{k,j}\right]_{b,a} = \boldsymbol{0}$ if $b \neq L_k$.

$T3$**:** Only transitions from ω_k to ω_j, i.e., $\boldsymbol{Q}_{k,j} = \boldsymbol{0}$.

$T4$**:** Only transitions from ω_j to ω_k (Reverse $T3$ transition), i.e., $\boldsymbol{Q}_{j,k} = \boldsymbol{0}$.

$T5$: No transitions between ω_j and ω_k, i.e., $\boldsymbol{Q}_{j,k} = \boldsymbol{0}$ and $\boldsymbol{Q}_{k,j} = \boldsymbol{0}$. ■

Note that $T1$ and $T2$ are mutually exclusive, except for the trivial case of all zero, but that other sets may have a non-empty intersection, for example $T1$ and $T3$, and $T2$ and $T3$, etc.

These five sets of transitions are shown in an example in Figure 1. In this small example we consider a network of connected QBD-processes and focus on ω_i and ω_j, each with 4 levels, and their one-step transitions. For each of the five sets of transitions from Definition 1 we present a schematic view of the generator. In this schematic view we depict a (possibly) non-zero submatrix by a gray square. The dark gray squares depict the direct connections between ω_i and ω_j. The white squares depict zero-submatrices. In Figure 1 it is easy to see that in a $T1$ transition there are transitions from any level in ω_i to l_1 in ω_j and back. A $T4$ transitions shows that there are only transitions from ω_j to ω_i. Figure 1 makes it easy to visualise how the intersection of $T1$ and $T4$, with transitions from l_1 in ω_j to any level in ω_i but none back, looks like. Finally, observe that a $T5$ transition is the trivial all-zero intersection of $T1, \ldots, T4$.

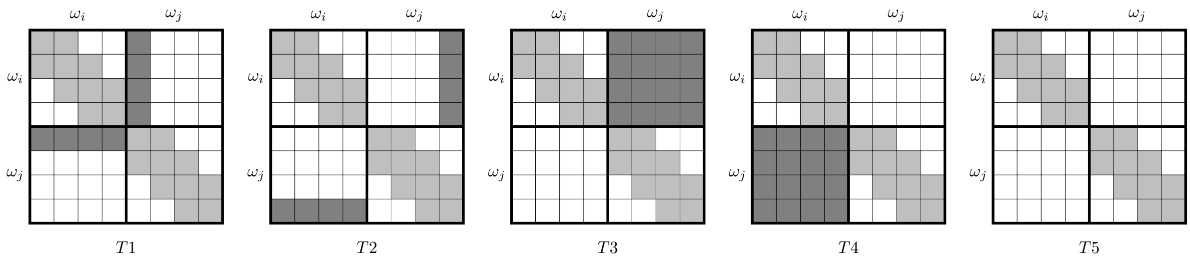

Figure 1: Schematic representation of the generators corresponding to each of the five types of transitions between ω_i and ω_j.

Definition 2. Indirect conditions describe the multiple step paths between ω_j and ω_k. We define a **lower path** from ω_j to ω_k as a path from ω_j to ω_k only passing through sets with index less than $\max\{j,k\}$. Based on the one step transitions between ω_j and ω_k, $(j < k)$ in Definition 1, we define the following indirect conditions:

i. If there is a $T1$ transition then:

- **a.** Each lower path from ω_j to ω_k must end with a $T1$ transition, and,
- **b.** Each lower path from ω_k to ω_j must start with a $T1$ transition.

ii. If there is a $T2$ transition then:

- **a.** Each lower path from ω_j to ω_k must end with a $T2$ transition, and,
- **b.** Each lower path from ω_k to ω_j must start with a $T2$ transition.

iii. If there is a $T3$ transition then there cannot be a lower path from ω_k to ω_j.

iv. If there is a $T4$ transition then there cannot be a lower path from ω_j to ω_k.

v. If there is a $T5$ transition then either:

a. All lower paths from ω_k to ω_j start with a $T1$ transition and all lower paths from ω_j to ω_k end with a $T1$ transition, or,

b. All lower paths from ω_k to ω_j start with a $T2$ transition and all lower paths from ω_j to ω_k end with a $T2$ transition, or,

c. There can be one or more lower paths from ω_j to ω_k, but none from ω_k to ω_j, or,

d. There can be one or more lower paths from ω_k to ω_j, but none from ω_j to ω_k, or,

e. There are no lower paths between ω_j and ω_k. ■

Remark 1 (Difference between transition types.)**.** In Figure 1 it appears that there is no difference between $T1$ and $T2$ transitions, since one can easily reorder the levels of ω_j in decreasing order, and a $T1$ ($T2$) transition becomes a $T2$ ($T1$) transition. However, in the example in Figure 2, in which ω_k with 4 levels is added, it is clear that it is not always possible to remove a $T2$ transition by reordering the levels in a certain set.

Also, it appears that there is no difference between $T3$ and $T4$ transitions. By interchanging ω_i and ω_j $T3$ ($T4$) transitions become $T4$ ($T3$) transitions. In Section 5.2 we show that for any Markov chain of connected QBD-processes with both $T3$ and $T4$ transitions we can reorder the sets such that there are only $T3$ transitions, making the $T4$ transition redundant. Nevertheless, for sake of clarity in notation we will introduce and use a $T4$ transition in our successive censoring algorithm.

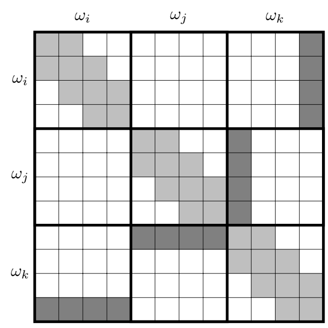

Figure 2: Schematic representation of a generator with three sets and both a $T1$ and a $T2$ transition.

In Section 4 we present an algorithm which identifies all sets and determines an ordering such that successive censoring algorithm of Section 3 can be applied.

Remark 2 (Special cases)**.** We will briefly discuss the relation between our model and the models discussed in Katehakis and Smit [9] and Gaver, Jacobs and Latouche [6]. In Gaver, Jacobs and Latouche

[6], a successive censoring algorithm is presented to find the stationary distribution of a Level-Dependent Quasi-Birth-and-Death process (LDQBD). By assuming that each set consists of a single level, and by assuming that there are only transition from ω_j to ω_{j+1} and back, $j = 1, \dots, S-1$, we obtain a LDQBD-process. In this special case, our successive censoring algorithm is the same as the successive censoring algorithm of Gaver, Jacobs and Latouche [6].

In Katehakis and Smit [9] a successive lumping procedure is presented for a special class of Markov Chains. Important is that the state space can be partitioned into sets and that in each set there is only one single entrance state, a state through which the set is entered. Note that there are no restrictions for the transitions within a set. By assuming that all levels consists of a single phase, and by restricting to $T1$ and $T2$ transitions only, we obtain a special case of both our model and the model by Katehakis and Smit [9].

3 Successive Censoring in detail

Let $\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 & \cdots & \boldsymbol{\pi}_S \end{bmatrix}$ denote the stationary distribution of the Markov Chain such that $\boldsymbol{\pi}\boldsymbol{Q} = \boldsymbol{0}$ and $\boldsymbol{\pi}\boldsymbol{e} = 1$ and let $\boldsymbol{\pi}_i$ denote the stationary distribution of $\omega_i, i \in \{1, \dots, S\}$. We obtain $\boldsymbol{\pi}$ by using a successive censoring algorithm based on the censoring algorithm in Kemeny and Snell [11] in Appendix A. In the censoring algorithm the state space of an arbitrary Markov Chain $\mathcal{Y}$ is first split into subsets A and B such that its generator $\boldsymbol{T}$ and stationary distribution ν can be partitioned following:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_A & \boldsymbol{T}_{AB} \\ \boldsymbol{T}_{BA} & \boldsymbol{T}_B \end{bmatrix}, \qquad \nu = \begin{bmatrix} \nu_A & \nu_B \end{bmatrix}.$$

Then a reduction step occurs in which transitions from B to B via A are projected onto transitions within B creating the generator $\boldsymbol{T}_B^*$:

$$\boldsymbol{T}_B^* = \boldsymbol{T}_B + \boldsymbol{T}_{BA} \left[-\boldsymbol{T}_A\right]^{-1} \boldsymbol{T}_{AB} \tag{2}$$

During an intermediate step the stationary distribution ν_B is determined by solving:

$$\nu_B \boldsymbol{T}_B^* = \boldsymbol{0},$$

and is used in the expansion step the determine ν_A:

$$\nu_A = \nu_B \boldsymbol{T}_{BA} \left[-\boldsymbol{T}_A\right]^{-1} \tag{3}$$

The successive censoring algorithm consists of $S-1$ reduction steps (2), an intermediate step, and $S-1$ expansion steps (3). In reduction step k, $k=1,\ldots,S-1$, the generator $\boldsymbol{Q}^k$ is reduced to $\boldsymbol{Q}^{k+1}$ by removing ω_k from the state space (censoring). Observe that following this definition, $\boldsymbol{Q}^1 = \boldsymbol{Q}$. In the intermediate step, the stationary distribution of $\boldsymbol{Q}^S$ is determined. Next, in expansion step k, $k=1,\ldots,S-1$, the stationary distribution is expanded by adding ω_{S-k}, the set with highest index still censored, back to the state space. Finally, by normalising the resulting vector, we obtain the stationary distribution $\boldsymbol{\pi}$.

Each $\boldsymbol{Q}_{i,i}$, $i=1,\ldots,S$, describes a transient Quasi-Birth-and-Death process and due to the irreducibility assumption its negative inverse $\left[-\boldsymbol{Q}_{i,i}\right]^{-1}$ exists and describes the sojourn time in ω_i before transition to some other ω_j. Let us denote by $\left[-\boldsymbol{Q}_{i,i}\right]^{-1}_{a,b}$, $a=1\ldots,L_j$ and $b=1\ldots,L_k$, the submatrix of $\left[-\boldsymbol{Q}_{i,i}\right]^{-1}$ describing the average time spent in l_b in ω_i before the Markov process leaves ω_i, given that it entered ω_i through l_a.

3.1 Reduction step k

In reduction step k, the generator $\boldsymbol{Q}^k$ is reduced to $\boldsymbol{Q}^{k+1}$ by removing ω_k from the state space. Observe that ω_k is the set with smallest index in $\boldsymbol{Q}^k$. Following the reduction step (2) we obtain for $i,j>k$

$$\boldsymbol{Q}^{k+1}_{i,j} = \boldsymbol{Q}^{k}_{i,j} + \boldsymbol{Q}^{k}_{i,k}\left[-\boldsymbol{Q}^{k}_{k,k}\right]^{-1}\boldsymbol{Q}^{k}_{k,j}.$$

Decomposing these submatrices by their levels, for $i=j>k$, gives:

$$\left[\boldsymbol{Q}^{k+1}_{i,i}\right]_{x,y} = \left[\boldsymbol{Q}^{k}_{i,i}\right]_{x,y} + \sum_{a=1}^{L_k}\sum_{b=1}^{L_k}\left[\boldsymbol{Q}^{k}_{i,k}\right]_{x,a}\left[-\boldsymbol{Q}^{k}_{k,k}\right]^{-1}_{a,b}\left[\boldsymbol{Q}^{k}_{k,i}\right]_{b,y}. \tag{4}$$

In this reduction step transitions from ω_i to ω_i via ω_k are projected onto transitions within ω_i. For example, a $T1$ transition from ω_k to ω_i is projected onto transitions within l_1 of ω_i and $x=y=1$ in (4). We rewrite (4) as:

$$\left[\boldsymbol{Q}^{k+1}_{i,i}\right]_{x,y} = \begin{cases} \left[\boldsymbol{Q}^{k}_{i,i}\right]_{x,y} + \sum_{a=1}^{L_k}\sum_{b=1}^{L_k}\left[\boldsymbol{Q}^{k}_{i,k}\right]_{x,a}\left[-\boldsymbol{Q}^{k}_{k,k}\right]^{-1}_{a,b}\left[\boldsymbol{Q}^{k}_{k,i}\right]_{b,y}, & \text{if } x=y=r \\ \left[\boldsymbol{Q}^{k}_{i,i}\right]_{x,y}, & \text{otherwise.} \end{cases} \tag{5}$$

Here, r depends on the type of transition between ω_k and ω_i and is given in Table 1. When $r=1$ all transitions between ω_k and ω_i are projected onto l_1 of ω_i. Observe that $T3$, $T4$ and $T5$ transitions are not projected onto transitions within ω_i since there are no transitions from ω_i to ω_i via ω_k and

$$\left[\boldsymbol{Q}_{i,i}^{k+1}\right]_{x,y} = \left[\boldsymbol{Q}_{i,i}^{k}\right]_{x,y}.$$

We denote this by "-" in Table 1.

$T1$	$T2$	$T3$	$T4$	$T5$
$r = 1$	$r = L_i$	-	-	-

Table 1: Value of r for each type of transition from ω_k to ω_i $(k < i)$.

A similar decomposition as (4) applies for $k < i < j$:

$$\left[\boldsymbol{Q}_{i,j}^{k+1}\right]_{x,y} = \begin{cases} \left[\boldsymbol{Q}_{i,j}^{k}\right]_{x,y} + \sum_{a=1}^{L_k}\sum_{b=1}^{L_k}\left[\boldsymbol{Q}_{i,k}^{k}\right]_{x,a}\left[-\boldsymbol{Q}_{k,k}^{k}\right]_{a,b}^{-1}\left[\boldsymbol{Q}_{k,j}^{k}\right]_{b,y}, & \text{if } x = r_1, y = r_2 \\ \left[\boldsymbol{Q}_{i,j}^{k}\right]_{x,y}, & \text{otherwise,} \end{cases} \tag{6}$$

and

$$\left[\boldsymbol{Q}_{j,i}^{k+1}\right]_{x,y} = \begin{cases} \left[\boldsymbol{Q}_{j,i}^{k}\right]_{x,y} + \sum_{a=1}^{L_k}\sum_{b=1}^{L_k}\left[\boldsymbol{Q}_{j,k}^{k}\right]_{x,a}\left[-\boldsymbol{Q}_{k,k}^{k}\right]_{a,b}^{-1}\left[\boldsymbol{Q}_{k,i}^{k}\right]_{b,y}, & \text{if } x = s_1, y = s_2 \\ \left[\boldsymbol{Q}_{j,i}^{k}\right]_{x,y}, & \text{otherwise,} \end{cases} \tag{7}$$

The ranges r_1, r_2, s_1 and s_2 depend on the transitions between ω_k and ω_j and between ω_k and ω_i. For $i < j$ these ranges are given in Table 2. For example, suppose there are $T1$ transitions from ω_k to ω_i and $T3$ transitions from ω_k to ω_j $(i < j)$. In reduction step k, these transitions will be projected onto transitions from the l_1 in ω_i to any level in ω_j ($r_1 = 1$ and $r_2 = 1, \ldots, L_j$) and no transitions from ω_j to ω_i.

Note that during reduction step k the transition from (or via) ω_k are projected onto existing transitions between sets ω_x, $x > k$. Using this we can now formulate the following theorem relating the indirect regulations in Definition 2 to the direct regulations in Definition 1.

Theorem 1. *The indirect regulations in Definition 2 ensure that the direct regulations in Definition 1 are preserved in each reduction step.*

Proof. Observe that a lower path from ω_j to ω_k, $j < k$, is projected onto a direct transition from ω_j to ω_k in reduction steps $1, \ldots, j-1$. Therefore, following the order in Definition 2, we can easily state that:

i. The lower paths in Def. 2.**i.a.** and Def. 2.**i.b.** will be projected onto transitions from any level in ω_j to l_1 in ω_k and onto transitions from l_1 in ω_k to any level in ω_j, respectively, i.e., $\left[\boldsymbol{Q}_{j,k}^{j-1}\right]_{a,b} = \mathbf{0}$

Transition from ω_k to ω_i $(k < i)$		Transition from ω_k to ω_j $(k < j)$			
		$T1$	$T2$	$T3$	$T4$
	$T1$	$r_1 = 1$ $r_2 = 1$ $s_1 = 1$ $s_2 = 1$	$r_1 = 1$ $r_2 = L_j$ $s_1 = L_j$ $s_2 = 1$	$r_1 = 1$ $r_2 = 1, \ldots, L_j$ - -	- - $s_1 = 1, \ldots, L_j$ $s_2 = 1$
	$T2$	$r_1 = L_i$ $r_2 = 1$ $s_1 = 1$ $s_2 = L_i$	$r_1 = L_i$ $r_2 = L_j$ $s_1 = L_j$ $s_2 = L_i$	$r_1 = L_i$ $r_2 = 1, \ldots, L_j$ - -	- - $s_1 = 1, \ldots, L_j$ $s_2 = L_i$
	$T3$	- - $s_1 = 1$ $s_2 = 1, \ldots, L_i$	- - $s_1 = L_j$ $s_2 = 1, \ldots, L_i$	- - - -	- - $s_1 = 1, \ldots, L_j$ $s_2 = 1, \ldots, L_i$
	$T4$	$r_1 = 1, \ldots, L_i$ $r_2 = 1$ - -	$r_1 = 1, \ldots, L_i$ $r_2 = L_j$ - -	$r_1 = 1, \ldots, L_i$ $r_2 = 1, \ldots, L_j$ - -	- - - -

Table 2: Ranges r_1, r_2, s_1 and s_2 for different types of transition between ω_k and ω_j and between ω_k and ω_i for $k < i < j$.

and $\left[\boldsymbol{Q}_{k,j}^{j-1}\right]_{b,a} = \mathbf{0}$ if $b \neq 1$ thus preserving the $T1$ transitions.

ii. The lower paths in Def. 2.**ii.a.** and Def. 2.**ii.b.** will be projected onto transitions from any level in ω_j to l_{L_k} in ω_k and onto transitions from l_{L_k} in ω_k to any level in ω_j, respectively, i.e. $\left[\boldsymbol{Q}_{j,k}^{j-1}\right]_{a,b} = \mathbf{0} and \left[\boldsymbol{Q}_{k,j}^{j-1}\right]_{b,a} = \mathbf{0}$ if $b \neq L_k$ thus preserving the $T2$ transitions.

iii. There are no lower paths from ω_k to ω_j so $\boldsymbol{Q}_{k,j}^{j-1} = \mathbf{0}$ and the $T3$ transitions are preserved.

iv. There are no lower paths from ω_j to ω_k so $\boldsymbol{Q}_{j,k}^{j-1} = \mathbf{0}$ and the $T4$ transitions are preserved.

v. Following the above reasoning we immediately state that the lower paths are projected onto:

a. a $T1$ transition.

b. a $T2$ transition.

c. a $T3$ transition.

d. a $T4$ transition.

e. a $T5$ transition.

Since $T5$ transitions can be considered as special cases of $T1$, $T2$, $T3$ and $T4$ transitions we can conclude that the direct regulations are maintained in each reduction step by the indirect regulations. □

Theorem 1 ensures that the fives types of transitions in Definition 1 are maintained through all the reduction steps. We can therefore state the following relation between the direct regulations and the QBD-structure of each set.

Theorem 2. *The direct regulations between ω_j and ω_k, $j < k$, in Definition 1 ensure that the original QBD-structure of ω_k is preserved in reduction step j. Furthermore, these five types of transitions are the only transitions that preserve the QBD-structure.*

Proof. From (4) and Table 1 it can be seen that both $T1$ and $T2$ transitions are projected onto transitions within the boundary levels of the QBD-process, namely the first level for a $T1$ transition and the last level for a $T2$ transition. It also follows from Table 1 that the remaining three types of transitions are not projected onto the QBD-process and we conclude that the direct regulations preserve the QBD-structure of ω_k.

Now consider a $T6$ transition different from the fives types in Definition 1. If there are transitions in only one direction a $T6$ transition is merely a special case of a $T3$ or a $T4$ transition, so suppose there are transitions in both directions. Note that since ω_j is removed from the state space before ω_k, it does not matter which levels in ω_j these transitions are going to or coming from. Next, note that, to preserve the QBD-structure, transitions can only be projected onto transitions within the first or the last level in ω_k. This means that if there are transitions from any level in ω_j to more than one level in ω_k, there cannot be any transitions from ω_k to ω_j ($T3$), else the QBD-structure no longer exists. Similarly, if there are transitions from more than one level in ω_k to any level in ω_j, there cannot be any transitions from ω_j to ω_k ($T4$). So finally suppose that there are transitions from any level in ω_j to level a in ω_k and transitions from level b in ω_k to any level in ω_j. Such transitions will be projected to direct transitions from level a to b within ω_k and will only preserve the QBD-structure if $a = b = 1$ ($T1$) or $a = b = L_k$ ($T2$). We can thus conclude that the five types of transitions in Definition 1 are the only types that preserve the QBD-structure of ω_k. □

Theorem 2 guarantees that $\boldsymbol{Q}_{i,i}^k$ describes a QBD-process for $i = k, \ldots, S$.

3.2 Intermediate step

From Theorem 2 we can conclude that $\boldsymbol{Q}^S$ describes a finite QBD-process of L_S levels:

$$\boldsymbol{Q}^S = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{F} & & & \\ \boldsymbol{B} & \boldsymbol{L} & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \boldsymbol{L} & \boldsymbol{F} \\ & & & \boldsymbol{B} & \boldsymbol{Y} \end{bmatrix}.$$

The stationary distribution $\boldsymbol{p}_S = \begin{bmatrix} \boldsymbol{p}_S^1 & \boldsymbol{p}_S^2 & \cdots & \boldsymbol{p}_S^{L_S} \end{bmatrix}$ of $\boldsymbol{Q}^S$ is given by, see Theorem 10.3.2 in Latouche and Ramaswami [14]:

$$\boldsymbol{p}_S^k = x_0 \boldsymbol{R}_1^{k-1} + x_1 \boldsymbol{R}_2^{L_S-k}, \tag{8}$$

where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are the minimal non-negative solutions to

$$\boldsymbol{F} + \boldsymbol{R}_1 \boldsymbol{L} + \boldsymbol{R}_1^2 \boldsymbol{B} = \boldsymbol{0} \qquad\qquad \boldsymbol{R}_2^2 \boldsymbol{F} + \boldsymbol{R}_2 \boldsymbol{L} + \boldsymbol{B} = \boldsymbol{0},$$

and $\begin{bmatrix} x_0 & x_1 \end{bmatrix}$ is the solution of the system

$$\begin{bmatrix} x_0 & x_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X} + \boldsymbol{R}_1 \boldsymbol{B} & \boldsymbol{R}_1^{L_S-2} [\boldsymbol{F} + \boldsymbol{R}_1 \boldsymbol{Y}] \\ \boldsymbol{R}_2^{L_S-2} [\boldsymbol{R}_2 \boldsymbol{X} + \boldsymbol{B}] & \boldsymbol{R}_2 \boldsymbol{F} + \boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \end{bmatrix},$$

and

$$x_0 \sum_{i=0}^{L_S-1} \boldsymbol{R}_1^i \boldsymbol{e} + x_1 \sum_{i=0}^{L_S-1} \boldsymbol{R}_2^i \boldsymbol{e} = 1,$$

where $\boldsymbol{e}$ denotes vectors of ones of an appropriate size.

3.3 Expansion step

Let $\begin{bmatrix} \boldsymbol{p}_{S-k} & \cdots & \boldsymbol{p}_{S-1} & \boldsymbol{p}_S \end{bmatrix}$ be the vector obtained after expansion step k. By normalising this vector we obtain the stationary distribution of $\boldsymbol{Q}^{S-k}$. Let $[\boldsymbol{p}_i]_j$ denote the subvector of $\boldsymbol{p}_i$ corresponding to level j in ω_i. Following the expansion step (3) we obtain:

$$\begin{aligned}
\boldsymbol{p}_{S-k} &= \begin{bmatrix} \boldsymbol{p}_{S-k+1} & \cdots & \boldsymbol{p}_S \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}^{S-k}_{S-k+1,S-n} \\ \vdots \\ \boldsymbol{Q}^{S-n}_{S,S-k} \end{bmatrix} \left[-\boldsymbol{Q}^{S-k}_{S-k,S-k}\right]^{-1} \\
&= \sum_{i=1}^{k} \boldsymbol{p}_{S-k+i} \boldsymbol{Q}^{S-k}_{S-k+i,S-k} \left[-\boldsymbol{Q}^{S-k}_{S-k,S-k}\right]^{-1}
\end{aligned}$$

By decomposing the submatrices by their levels gives:

$$\left[\boldsymbol{p}_{S-k}\right]_j = \sum_{i=1}^{k} \sum_{b=1}^{L_{S-k}} \sum_{a=1}^{L_{S-k+i}} \left[\boldsymbol{p}_{S-k+i}\right]_a \left[\boldsymbol{Q}^{S-k}_{S-k+i,S-k}\right]_{a,b} \left[-\boldsymbol{Q}^{S-k}_{S-k,S-k}\right]^{-1}_{b,j} \tag{9}$$

By utilising the type of transition between ω_{S-k+i} and ω_{S-k} we can write the inner sum as

$$\sum_{a=t_1}^{t_2} \left[\boldsymbol{p}_{S-k+i}\right]_a \left[\boldsymbol{Q}^{S-k}_{S-k+i,S-k}\right]_{a,b}.$$

The values of t_1 and t_2 follow from the type of the transition and are given in Table 3.

$T1$	$T2$	$T3$	$T4$	$T5$
$t_1 = 1$	$t_1 = L_{S-k+i}$	-	$t_1 = 1$	-
$t_2 = 1$	$t_2 = L_{S-k+i}$	-	$t_2 = L_{S-k+i}$	-

Table 3: Ranges t_1 and t_2 for each type of transition from ω_{S-k} to ω_{S-k+i}.

The stationary distribution $\boldsymbol{\pi}$ of $\boldsymbol{Q}$ is obtained by normalising the vector obtained after expansion step $S-1$.

3.4 Inverse of $-\boldsymbol{Q}^k_{k,k}$

In reduction step k and expansion step $S-k$ the negative inverse of the transient generator $\boldsymbol{Q}^k_{k,k}$ need to be determined. It follows from Theorem 1 and Theorem 2 that $\boldsymbol{Q}^k_{k,k}$ describes a QBD-process for $k = 1, \ldots, S$. In Choi et al [4] direct formulas are given to determine the fundamental matrix of a transient QBD, see Appendix B, which rely on determining $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, the minimal non-negative solutions to

$$\boldsymbol{F} + \boldsymbol{R}_1\boldsymbol{L} + \boldsymbol{R}_1^2\boldsymbol{B} = \boldsymbol{0} \qquad \boldsymbol{R}_2^2\boldsymbol{F} + \boldsymbol{R}_2\boldsymbol{L} + \boldsymbol{B} = \boldsymbol{0}.$$

Here $\boldsymbol{F}$, $\boldsymbol{L}$ and $\boldsymbol{B}$ are the forward (transition from level i to $i+1$), local (transitions within level i) and backward (transitions from level $i+1$ to i) transition matrices describing the QBD-process. These matrix equations can easily be solved by the fixed-point iteration in Neuts [17] or the efficient logarithmic reduction algorithm by Latouche and Ramaswami [13].

Note that the direct formulas of Choi et al. in Appendix B require the inverse of a $4P_k \times 4P_k$ matrix for ω_k, therefore, if $L_k \leq 4$ it is beneficial to determine the inverse of $-\boldsymbol{Q}^k_{k,k}$ directly instead of using the direct formulas of Choi et al.

Remark 3 (Infinite sized sets)**.** In this paper we assume that all the sets have a finite size but this is not necessary. By carefully choosing the transitions between sets it becomes possible to have an infinite number of levels in each set. For example, if ω_S has an infinite number of levels, we must determine the stationary distribution of a regular QBD-process in the intermediate step. Also, further results by Choi et al [4] include direct formulas for the fundamental matrix of a transient QBD with infinite levels.

The $T2$ transitions are meaningless for infinite sized sets and according to Corollary 1, $T4$ transitions can be transformed to $T3$ transitions. Furthermore, if there are $T1$ or $T3$ transitions, there must be a functional relationship between the submatrices involved in the equations (5), (6) and (7), such that these equations can still be computed.

4 Ordering of the sets

In Section 2 we introduced a system of connected QBD-processes for which a successive censoring algorithm was introduced in Section 3. An important part of this algorithm is the ordering of the sets (QBD-processes), which was implicitly mentioned in Definition 1 through the direction of the transitions, and in Definition 2 through the notion of a **lower path**.

In this section we will introduce an algorithm which determines whether the stationary distribution of a given Markov Chain can be obtained with the successive censoring algorithm. As a bonus, it presents both the sets and the ordering of the sets needed in the successive censoring algorithm.

Algorithm 1.

Step 1 Identify the sets ω_j such that each ω_j is a QBD-process.

Step 2 Check, for each pair of sets ω_j and ω_k, regardless of their order ($j < k$ or $j > k$), if the direct conditions in Definition 1 hold.

Step 3 Determine the ordering $\boldsymbol{\sigma} = \{\sigma(1), \ldots, \sigma(S)\}$ such that also the direction of the direct conditions hold. If there are transitions from any level in ω_k to the first (last) level in ω_j and back, then $\sigma(\omega_k) < \sigma(\omega_j)$. Also, if there are transitions from any level in ω_k to any level in ω_j, by none back, $\sigma(\omega_k) < \sigma(\omega_j)$.

Step 4 Finally, check, for the ordered system of connected QBD-processes if the indirect conditions in Definition 2 hold.

If, for a given Markov Chain, all four steps can be successfully executed, it is possible to obtain the stationary distribution for this Markov Chain with the help of the successive censoring algorithm. This is not difficult to see since in **Step 1** we check if the state space $\mathcal{S}$ can be partitioned into sets each describing a QBD-process, in **Step 2** and **3** we check if there is an ordering such that the direct condition are met, and in **Step 4** we check if, for this ordering, the indirect conditions are met.

Observe that, since we do not have a lower bound on the number of levels in a QBD-process, **Step 1** can be executed for any Markov Chain. Also observe that this partition is not unique. Suppose for some partition there are two sets ω_i and ω_j such that there are transitions from any level in ω_i to some l_x, with $x \neq 1, L_j$, in ω_j and back. Then the direct conditions in Definition 1 are not met. However, by correctly splitting up ω_j into sets ω_{j_a} and ω_{j_b} we obtain a new partition such that the direct conditions are met.

For a given Markov Chain with its state space partitioned according to **Step 1** it is not difficult to order the sets and check whether the direct conditions in Definition 1 hold. It is only necessary to focus on $T1$ and $T2$ transitions and make sure that have the correct orientation. However, it is a demanding task to verify if the indirect conditions in Definition 2 hold, since all lower paths from ω_i to ω_j (and back) need to be considered. We solve this problem by using a simplified version of the successive censoring algorithm. Let $\mathcal{C}^k(i,j)$, $i<j$, denote the collection of transition types from ω_i to ω_j after reduction step $k-1$ (in which ω_{k-1} is removed), i.e., if $\mathcal{C}^k(i,j) = \{T1, T3\}$, l_1 of ω_j can be reached from any level in ω_i creating a special case of both $T1$ and $T3$ transitions. If there are $T5$ transitions from ω_i to ω_j we state $\mathcal{C}^k(i,j) = \{T1, T2, T3, T4\}$.

Next, we define the iteration

$$\mathcal{C}^{k+1}(i,j) = C^k(i,j) \cap \mathcal{W}\left[\mathcal{C}^k(k,i) \times \mathcal{C}^k(k,j)\right], \qquad i<j \tag{10}$$

where the Cartesian product $\mathcal{C}^k(k,i) \times \mathcal{C}^k(k,j)$ consists of all ordered pairs describing the type of transitions from ω_k to ω_i and from ω_k to ω_j. Upon removing ω_k these transitions are projected to transitions from ω_i to ω_j according to Table 4. The function $\mathcal{W}$ is then the intersection of the projections of each pair in $\mathcal{C}^k(k,i) \times \mathcal{C}^k(k,j)$.

		$\mathcal{C}^k(k,j),\ k<j$			
		$T1$	$T2$	$T3$	$T4$
$\mathcal{C}^k(k,i),\ k<i$	$T1$	$T1$	$T2$	$T3$	$T4$
	$T2$	$T1$	$T2$	$T3$	$T4$
	$T3$	$\{T1, T4\}$	$\{T2, T4\}$	$\{T1, T2, T3, T4\}$	$T4$
	$T4$	$\{T1, T3\}$	$\{T2, T3\}$	$T3$	$\{T1, T2, T3, T4\}$

Table 4: Projections onto transitions from ω_i to ω_j, $k<i<j$.

For example, suppose $\mathcal{C}^k(i,j)=\{T1,T3\}$, $\mathcal{C}^k(k,i)=\{T2,T3\}$ and $\mathcal{C}^k(k,j)=\{T1,T4\}$ then:

$$\begin{aligned}\mathcal{W}\left[\mathcal{C}^k(k,i)\times\mathcal{C}^k(k,j)\right] &= \mathcal{W}\left[\{T2,T1\},\{T2,T4\},\{T3,T1\},\{T3,T4\}\right]\\ &= T1\cup T4\cup\{T1,T4\}\cup T4\\ &= \{T1,T4\}\end{aligned}$$

and

$$\mathcal{C}^{k+1}(i,j)=\{T1,T3\}\cap\{T1,T4\}=T1$$

Theorem 3. *If $\mathcal{C}^k(i,j)=\emptyset$ for any two sets ω_i and ω_j, $i<j$, after reduction step k ($k=1,\ldots,\omega-1$), then the direct regulations in Definition 1 are violated and the successive censoring algorithm can no longer be applied.*

Proof. Observe that the first term in (10) describes the possible types of transitions from ω_i to ω_j before removing ω_k, whereas the second term describes the projection as a result of removing ω_k. If this projection is different than the existing types of transitions, $\mathcal{C}^{k+1}(i,j)=\emptyset$ and the direct regulations are violated after reduction step k. □

5 The successive censoring algorithm

In this section we will give a short overview of the complete successive censoring algorithm, demonstrate the algorithm with an example and determine its complexity.

5.1 The full algorithm

The complete successive censoring algorithm is summarised by the following algorithm

Algorithm 2 (The successive censoring algorithm)**.**

1. Determine, for a given Markov Chain $\mathcal{X}$. if the successive censoring algorithm can be applied using Algorithm 1. Also, using Algorithm 1, identify the sets, determine the number of sets, S, and determine their ordering.

2. Reduce the generator in $S-1$ reduction steps using equations (5), (6) and (7) in Section 3.1.

3. Determine the stationary distribution of $\boldsymbol{Q}^S$ using (8) in Section 3.2.

4. Expand the stationary distribution of $\boldsymbol{Q}^S$ in $S-1$ expansion steps using equation (9) in Section 3.3.

5. Finally, normalise the resulting vector to obtain the stationary distribution of the Markov Chain $\mathcal{X}$.

We demonstrate the successive censoring algorithm with an example based on the threshold queues by Baer, Boucherie and van Ommeren [2].

Example 1. Let us consider a single server queue, with buffer size N, in which service rates are controlled by a threshold policy. Customers arrive according to a Poisson process with rate λ and require an exponential service time, depending on the stage of the queue. When the queue is in stage 1, the service rate is μ_1, and when the queue is in stage 2, the service rate is μ_2. Transition between the two stages is controlled by the threshold policy given by a lower threshold, L, and an upper threshold, U. The stage changes from 1 to 2 when an arrival occurs when the queue length is U. The stage changes back from 2 to 1 when a departure occurs when the queue length is L. The state diagram for the threshold queue with 2 stages is given in Figure 3.

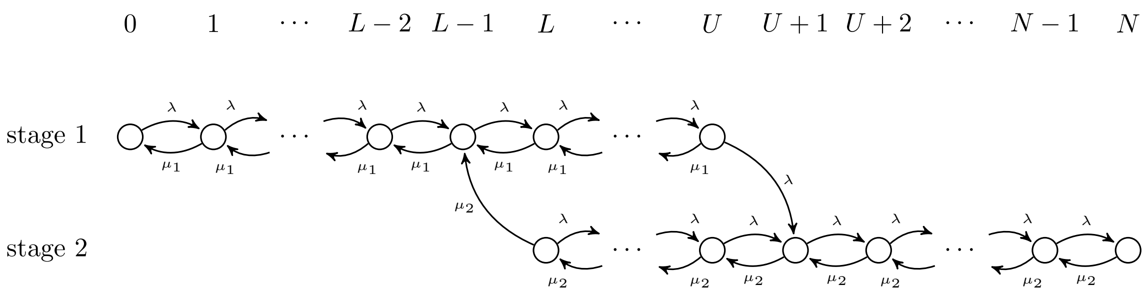

Figure 3: State diagram for the 2-stage threshold $M/M/1$ queue.

In this example we will consider a 2-stage threshold $M/M/1$ queue with $\lambda = 4$, $\mu_1 = 8$, $\mu_2 = 6$,

$L = 3$, $U = 6$, and $N = 10$, and with generator:

$$
\boldsymbol{Q} = \left[\begin{array}{ccccccc|cccc|cccc}
-4 & 4 & & & & & & & & & & & & & \\
8 & -12 & 4 & & & & & & & & & & & & \\
 & 8 & -12 & 4 & & & & & & & & & & & \\
 & & 8 & -12 & 4 & & & & & & & & & & \\
 & & & 8 & -12 & 4 & & & & & & & & & \\
 & & & & 8 & -12 & 4 & & & & & & & & \\
 & & & & & 8 & -12 & & & & & 4 & & & \\
\hline
 & & 6 & & & & & -10 & 4 & & & & & & \\
 & & & & & & & 6 & -10 & 4 & & & & & \\
 & & & & & & & & 6 & -10 & 4 & & & & \\
 & & & & & & & & & 6 & -10 & 4 & & & \\
\hline
 & & & & & & & & & & 6 & -10 & 4 & & \\
 & & & & & & & & & & & 6 & -10 & 4 & \\
 & & & & & & & & & & & & 6 & -10 & 4 \\
 & & & & & & & & & & & & & 6 & -6
\end{array}\right].
$$

Here, the solid lines denote three distinctive sets, each representing a QBD-process such that $\boldsymbol{Q}$ is represented by

$$
\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{1,1} & \boldsymbol{Q}_{1,2} & \boldsymbol{Q}_{1,3} \\ \boldsymbol{Q}_{2,1} & \boldsymbol{Q}_{2,2} & \boldsymbol{Q}_{2,3} \\ \boldsymbol{Q}_{3,1} & \boldsymbol{Q}_{3,2} & \boldsymbol{Q}_{3,3} \end{bmatrix}
$$

We will now apply Algorithm 2 to obtain the stationary distribution for this Markov Chain.

1. As can be seen above, the Markov Chain consists of three sets, depicted by the solid lines in the generator above. We will use the notation of Section 4 to denote the transitions between the sets and determine if the successive censoring algorithm can be applied:

$$
\mathcal{C}^1(1,2) = \{T1, T4\}, \qquad \mathcal{C}^1(1,3) = \{T1, T3\}, \qquad \mathcal{C}^1(2,3) = \{T1\},
$$

   end

$$
\begin{aligned}
\mathcal{C}^2(2,3) &= \mathcal{C}^1(2,3) \cap \mathcal{W}[\mathcal{C}^1(1,2) \times \mathcal{C}^1(1,3)] \\
&= \{T1\} \cap \{T1, T3\} = \{T1\}.
\end{aligned}
$$

As a result of Theorem 3 the successive censoring algorithm can be applied.

2. We will perform two consecutive reduction steps to reduce the generator. Since ω_1 can be reached via $T3$ and $T4$ transitions we first obtain:

$$\boldsymbol{Q}^2_{2,2} = \boldsymbol{Q}^1_{2,2}, \qquad \boldsymbol{Q}^2_{3,2} = \boldsymbol{Q}^1_{3,2}, \qquad \boldsymbol{Q}^2_{3,3} = \boldsymbol{Q}^1_{3,3},$$

and following (6) and using the results by Choi et al. in Appendix B:

$$\begin{aligned}
\left[\boldsymbol{Q}^2_{2,3}\right]_{1,1} &= \left[\boldsymbol{Q}^1_{2,3}\right]_{1,1} + \sum_{a=1}^{7}\sum_{b=1}^{7} \left[\boldsymbol{Q}^1_{2,1}\right]_{1,a} \left[-\boldsymbol{Q}^1_{1,1}\right]^{-1}_{a,b} \left[\boldsymbol{Q}^1_{1,3}\right]_{b,1} \\
&= \left[\boldsymbol{Q}^1_{2,3}\right]_{1,1} + \left[\boldsymbol{Q}^1_{2,1}\right]_{1,3} \left[-\boldsymbol{Q}^1_{1,1}\right]^{-1}_{3,7} \left[\boldsymbol{Q}^1_{1,3}\right]_{7,1} \\
&= 0 + 6 \cdot \frac{1}{4} \cdot 4 = 6 \\
\left[\boldsymbol{Q}^2_{2,3}\right]_{x,y} &= 0, \qquad\qquad x \neq 1, y \neq 1,
\end{aligned}$$

resulting in

$$\boldsymbol{Q}^2 = \left[\begin{array}{cccc|cccc}
-10 & 4 & & & 6 & & & \\
6 & -10 & 4 & & & & & \\
 & 6 & -10 & 4 & & & & \\
 & & 6 & -10 & 4 & & & \\
\hline
 & & & 6 & -10 & 4 & & \\
 & & & & 6 & -10 & 4 & \\
 & & & & & 6 & -10 & 4 \\
 & & & & & & 6 & -6
\end{array}\right].$$

There are $T1$ transitions from ω_2 to ω_3, therefore:

$$\begin{aligned}
\left[\boldsymbol{Q}^3_{3,3}\right]_{1,1} &= \left[\boldsymbol{Q}^2_{3,3}\right]_{1,1} + \left[\boldsymbol{Q}^2_{3,2}\right]_{1,4} \left[-\boldsymbol{Q}^2_{2,2}\right]_{4,1} \left[\boldsymbol{Q}^2_{2,3}\right]_{1,1} + \left[\boldsymbol{Q}^2_{3,2}\right]_{1,4} \left[-\boldsymbol{Q}^2_{2,2}\right]_{4,4} \left[\boldsymbol{Q}^2_{2,3}\right]_{4,1} \\
&= -10 + 6 \cdot \frac{27}{422} \cdot 6 + 6 \cdot \frac{65}{422} \cdot 4 = -4.
\end{aligned}$$

Which results in

$$\boldsymbol{Q}^3 = \begin{bmatrix}
-4 & 4 & & \\
6 & -10 & 4 & \\
 & 6 & -10 & 4 \\
 & & 6 & -6
\end{bmatrix}.$$

3. Next, we determine the stationary distribution of $\boldsymbol{Q}^3$ using the intermediate step in Section 3.2.

This results in:

$$\boldsymbol{R}_1 = \frac{2}{3}, \qquad \boldsymbol{R}_2 = 1, \qquad x_0 = \frac{27}{65}, \qquad x_1 = 0,$$

and

$$\boldsymbol{p}_3 = \begin{bmatrix} \frac{27}{65} & \frac{18}{65} & \frac{12}{65} & \frac{8}{65} \end{bmatrix}$$

4. We expand $\boldsymbol{p}_3$ with two consecutive expansion steps following (9):

$$\begin{aligned} [\boldsymbol{p}_2]_j &= \sum_{b=1}^{4}\sum_{a=1}^{4} [\boldsymbol{p}_3]_a \left[\boldsymbol{Q}^2_{3,2}\right]_{a,b} \left[-\boldsymbol{Q}^2_{2,2}\right]^{-1}_{b,j} \\ &= [\boldsymbol{p}_3]_1 \left[\boldsymbol{Q}^2_{3,2}\right]_{1,4} \left[-\boldsymbol{Q}^2_{2,2}\right]^{-1}_{4,j} \\ &= \frac{162}{65} \left[-\boldsymbol{Q}^2_{2,2}\right]^{-1}_{4,j}, \end{aligned}$$

which results in

$$\boldsymbol{p}_2 = \begin{bmatrix} \frac{2187}{13715} & \frac{729}{2743} & \frac{4617}{13715} & \frac{81}{211} \end{bmatrix}$$

Since ω_1 can only be reached from ω_2, the second expansion step gives

$$\begin{aligned} [\boldsymbol{p}_1]_j &= \sum_{i=1}^{2}\sum_{b=1}^{7}\sum_{a=1}^{L_{1+i}} \left[\boldsymbol{p}_{1+i}\right]_a \left[\boldsymbol{Q}^1_{1+i,1}\right]_{a,b} \left[-\boldsymbol{Q}^1_{1,1}\right]^{-1}_{b,j} \\ &= [\boldsymbol{p}_2]_1 \left[\boldsymbol{Q}^1_{2,1}\right]_{1,3} \left[-\boldsymbol{Q}^1_{1,1}\right]^{-1}_{3,j} \\ &= \frac{13122}{13715} \left[-\boldsymbol{Q}^1_{1,1}\right]^{-1}_{3,j} \end{aligned}$$

which results in

$$\boldsymbol{p}_1 = \begin{bmatrix} \frac{406782}{13715} & \frac{203391}{13715} & \frac{203391}{27430} & \frac{19683}{5486} & \frac{45927}{27430} & \frac{19683}{27430} & \frac{6561}{27430} \end{bmatrix}$$

5. Normalising the vectors $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ gives us the stationary distributions $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, and $\boldsymbol{\pi}_3$:

$$\begin{aligned} \boldsymbol{\pi}_1 &= \begin{bmatrix} \frac{813564}{1653181} & \frac{406782}{1653181} & \frac{203391}{1653181} & \frac{98415}{1653181} & \frac{45927}{1653181} & \frac{19683}{1653181} & \frac{6561}{1653181} \end{bmatrix} \\ \boldsymbol{\pi}_2 &= \begin{bmatrix} \frac{4374}{1653181} & \frac{7290}{1653181} & \frac{9234}{1653181} & \frac{10530}{1653181} \end{bmatrix} \\ \boldsymbol{\pi}_3 &= \begin{bmatrix} \frac{11394}{1653181} & \frac{7596}{1653181} & \frac{5064}{1653181} & \frac{3376}{1653181} \end{bmatrix} \end{aligned}$$

We can now check that the stationary distribution $\boldsymbol{\pi}$, such that $\boldsymbol{\pi Q} = \boldsymbol{0}$ and $\boldsymbol{\pi e} = 1$, is given by

$$\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 & \boldsymbol{\pi}_3 \end{bmatrix}$$

5.2 Complexity Analysis

Let us consider a system of connected QBD-processes with S sets, each with L $(= \max_k L_k)$ levels of P $(= \max_k P_k)$ phases each. The successive censoring algorithm consists of $S-1$ reduction steps, one intermediate step, and $S-1$ expansion steps. The complexity of the entire algorithm is determined by the complexity of the reduction steps. Since the intermediate step is performed only once while both the reduction and expansion steps are performed $S-1$ times, we can ignore the effect of the intermediate step on the complexity. Furthermore, some of the operations needed in the reduction steps are also needed in the expansions steps, the product $\boldsymbol{Q}^k_{i,k}[-\boldsymbol{Q}^k_{k,k}]^{-1}$ for example, is used in both the reduction step as well as the expansion step. However, in reduction step k, we need to multiply this product with a matrix (on the right) $(S-k)^2$ times, while in expansion step $S-k$ we must multiply this product with a vector (on the left) $S-k$ times. Therefore, the reduction step requires more computations than an expansion step and we can focus on the reduction steps alone to determine the complexity of the algorithm.

The complexity of each reduction step depends on the type of transitions between the sets. However, the worst-case scenario is a system of connected QBD-processes with $T1$ transition between all sets. To see this we first note that for the complexity of a reduction step there is no difference between a $T1$ or a $T2$ transition. Both types are projected with a vector-matrix-vector multiplication when combined with another $T1$ or $T2$ transition, and they are both projected with a vector-matrix-matrix or a matrix-matrix-vector multiplication when combined with a $T3$ or $T4$ transition respectively. Next, we show that by reordering the sets, any $T4$ transition can be changed into a $T3$ transition without disobeying the indirect conditions.

Corollary 1. *For any system of connected QBD-processes with both $T3$ and $T4$ transitions, the sets can be reordered such that there are only $T3$ transitions.*

Proof. Consider the (schematically represented) Markov Chain in Figure 4(a) with a $T4$ transition from ω_j to ω_i, $i < j$, depicted by the black arrow. Furthermore, let the blocks 1, 2, and 3 represent collections of sets with appropriate set-index (block 1 contains all ω_x with $x < i$, etc.). Definition 2.iv. specifies that there is no lower path from ω_i to ω_j via block 1 and/or block 2. This suggests that block 2 can be split up into two separate blocks 2a and 2b, such that block 2a contains sets connected to ω_i, and block 2b contains sets connected to ω_j. Note that because of Definition 2.iv. the only transitions between blocks 2a and 2b are $T4$ transitions from 2b to 2a. Therefore, we can reorder the sets as in Figure 4(b) in which the $T4$ transition from ω_j to ω_i is transformed into a $T3$ transition. Since there are no lower paths from ω_i to ω_j, Definition 2 holds and the successive censoring algorithm still applies. □

Due to Corollary 1 we will consider a system of connected QBD-processes with only $T1$ and $T3$

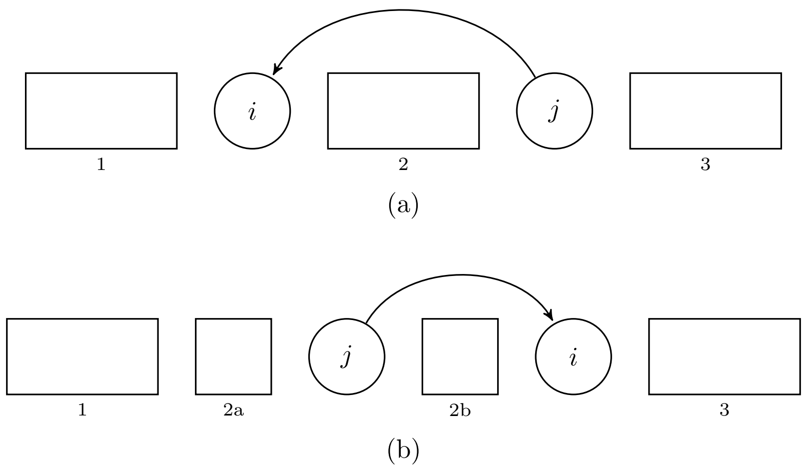

Figure 4: Schematic representation of system of connected QBD-processes, before (a) and after (b) reordering of the sets.

transitions. By only considering non-zero transitions, we can conclude that a projection of 2 $T1$ transitions is a vector-matrix-vector multiplication with $\mathcal{O}(L^2P^3)$, and that a projection of a $T1$ and a $T3$ transition is a vector-matrix-matrix multiplication, also with $\mathcal{O}(L^2P^3)$. So to determine the complexity of the algorithm we must maximise the number of projections made in each step (instead of the size of the projections). Since the projection of 2 $T1$ transitions results in 2 projections while the projection of a $T1$ and a $T3$ transition results in 1 projection, we will consider a system of connected QBD-processes with only $T1$ transitions.

In each reduction step k we must determine the fundamental matrix $[-Q^k_{k,k}]^{-1}$ using the results from Choi et al. The first step is to determine $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ in (16) which can be done with the logarithmic reduction algorithm with $\mathcal{O}(P^3)$ according to Latouche and Ramaswami [13]. Second, we determine $\boldsymbol{R}_1^k$ and $\boldsymbol{R}_2^k$ iteratively, taking $\mathcal{O}(LP^3)$ each. Using these matrices we create the matrices needed in (18) and (21), with $\mathcal{O}(P^3)$, and take their inverses, also with $\mathcal{O}(P^3)$. Finally, we determine the fundamental matrix using equations (17), (19) and (20). Observe that it takes $\mathcal{O}(P^3)$ to determine one submatrix. Since this operation must be performed L^2 times it gives $\mathcal{O}(L^2P^3)$. Due to this last, computationally heavy, step, the complexity of determining the fundamental matrix is $\mathcal{O}(L^2P^3)$.

Next, in reduction step, we must perform $(\omega-k)^2$ projections of 2 $T1$ transitions. A single projection of 2 $T1$ transitions is a vector-matrix-vector multiplication with $\mathcal{O}(L^2P^3)$. Therefore the total complexity, including the inverse following [4], of reduction step k is $\mathcal{O}(S^2L^2P^3)$. Finally, there are $\omega-1$ reduction steps resulting in a complexity of the successive censoring algorithm of $\mathcal{O}(S^3L^2P^3)$. Comparing this to solving $\pi\boldsymbol{Q}=\boldsymbol{0}$ with $\mathcal{O}(S^3L^3P^3)$ we conclude that we decrease complexity by a factor L.

6 Summary and Conclusion

We introduced a successive censoring algorithm to find the stationary distribution of a general class of Markov Chains consisting of multiple Quasi-Birth-and-Death processes (QBD) connected by special types of transitions. The successive censoring algorithm consists of reduction steps, in which the state space is reduced by removing a QBD, and expansion steps, in which the stationary distribution is expanded by adding a (previously removed) QBD. By applying the results of Choi et al [4] we determine the inverse of the transient QBD-generator required in both the reduction and expansion steps.

The successive censoring algorithm is summarised and applied to a 2-stage $M/M/1$ threshold queue in Section 5. Also it is shown that the complexity is $\mathcal{O}(S^3L^2P^3)$.

A Censoring Technique

Our successive censoring technique is based on the censoring technique in Kemeny and Snell for discrete time Markov Chains, see Chapter 6.1 in [11]. The extension to continuous time Markov Chains is described in Ye and Li [19].

Consider an irreducible continuous time Markov Chain on state space X. Partition X in subsets A and B so that $X = A \cup B$, $A \cap B = \emptyset$, and both $A \neq \emptyset$ and $B \neq \emptyset$. Let the generator $\boldsymbol{Q}$ be given by:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_A & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_B \end{bmatrix},$$

where $\boldsymbol{Q}_A$ and $\boldsymbol{Q}_B$ denote the transitions within A and B respectively and $\boldsymbol{Q}_{AB}$ and $\boldsymbol{Q}_{BA}$ denote the transitions between A and B.

In the reduction step the subset A is removed from X and the Markov Chain is observed on the subset B only. The reduced generator is given by:

$$\boldsymbol{Q}_B^* = \boldsymbol{Q}_B + \boldsymbol{Q}_{BA}\left[-\boldsymbol{Q}_A\right]^{-1}\boldsymbol{Q}_{AB}. \qquad \text{[reduction step]} \tag{11}$$

Let $\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_A & \boldsymbol{\pi}_B \end{bmatrix}$ be such that $\boldsymbol{\pi}\boldsymbol{Q} = \boldsymbol{0}$, then

$$\boldsymbol{\pi}_A\boldsymbol{Q}_A + \boldsymbol{\pi}_B\boldsymbol{Q}_{BA} = \boldsymbol{0} \tag{12}$$

$$\boldsymbol{\pi}_A\boldsymbol{Q}_{AB} + \boldsymbol{\pi}_B\boldsymbol{Q}_B = \boldsymbol{0} \tag{13}$$

Since the Markov Chain is irreducible, the inverse of $\boldsymbol{Q}_A$ exists and (12) gives

$$\boldsymbol{\pi}_A = \boldsymbol{\pi}_B \boldsymbol{Q}_{BA} \left[-\boldsymbol{Q}_A\right]^{-1}. \qquad \text{[expansion step]} \tag{14}$$

Inserting this in (13) gives

$$\boldsymbol{0} = \boldsymbol{\pi}_B \boldsymbol{Q}_B + \boldsymbol{\pi}_B \boldsymbol{Q}_{BA} \left[-\boldsymbol{Q}_A\right]^{-1} \boldsymbol{Q}_{AB} = \boldsymbol{\pi}_B \boldsymbol{Q}_B^*. \tag{15}$$

Once $\boldsymbol{\pi}_B$ is obtained from (15), $\boldsymbol{\pi}_A$ (and thus $\boldsymbol{\pi}$) is uniquely determined by the expansion step (3). By normalising $\boldsymbol{\pi}$ we obtain the stationary distribution of $\boldsymbol{Q}$.

B Inverse of transient QBD

The successive censoring algorithm requires the inverse of the generator of a transient QBD-process with M levels. This inverse is obtained by Choi et. al. in [4]. Let $\boldsymbol{Q}$ be the generator of a transient QBD:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{F} & & & \\ \boldsymbol{B} & \boldsymbol{L} & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \boldsymbol{L} & \boldsymbol{F} \\ & & & \boldsymbol{B} & \boldsymbol{Y} \end{bmatrix},$$

and let $\boldsymbol{e}$ denote a vector of all ones. Suppose $(\boldsymbol{B}+\boldsymbol{L}+\boldsymbol{F})$ is conservative (i.e. $(\boldsymbol{B}+\boldsymbol{L}+\boldsymbol{F})\boldsymbol{e} = \boldsymbol{0}$). Let κ denote its stationary distribution, $\kappa(\boldsymbol{B}+\boldsymbol{L}+\boldsymbol{F}) = \boldsymbol{0}$ and $\kappa\boldsymbol{e} = 1$, and let $\rho = (\kappa\boldsymbol{F}\boldsymbol{e})/(\kappa\boldsymbol{B}\boldsymbol{e})$. In this appendix we assume that $(\boldsymbol{B}+\boldsymbol{L}+\boldsymbol{F})$ is either transient or conservative with $\rho \neq 1$ and that $M < \infty$. We refer the reader to [4] for the case in which $(\boldsymbol{B}+\boldsymbol{L}+\boldsymbol{F})$ is conservative and $\rho = 1$ or the case where $M = \infty$.

We define $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ as the minimal non-negative solutions to

$$\boldsymbol{F} + \boldsymbol{R}_1\boldsymbol{L} + \boldsymbol{R}_1^2\boldsymbol{B} = \boldsymbol{0} \qquad \boldsymbol{R}_2^2\boldsymbol{F} + \boldsymbol{R}_2\boldsymbol{L} + \boldsymbol{B} = \boldsymbol{0}. \tag{16}$$

Finally, we denote by $\boldsymbol{Z}$ the inverse of $[-\boldsymbol{Q}]$:

$$[-\boldsymbol{Q}]^{-1} = \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}(1,1) & \boldsymbol{Z}(1,2) & \cdots & \boldsymbol{Z}(1,M) \\ \boldsymbol{Z}(2,1) & \boldsymbol{Z}(2,2) & \cdots & \boldsymbol{Z}(2,M) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Z}(M,1) & \boldsymbol{Z}(M,2) & \cdots & \boldsymbol{Z}(M,M) \end{bmatrix}$$

The rows of $\boldsymbol{Z}$ follow from Theorem 6 in [4]:

(i) The first and last rows are given by:

$$\begin{aligned} \boldsymbol{Z}(1,k) &= V(1,1)\boldsymbol{R}_1^{k-1} + V(1,2)\boldsymbol{R}_2^{M-k}, \quad 1 \le k \le M \\ \boldsymbol{Z}(M,k) &= V(2,1)\boldsymbol{R}_1^{k-1} + V(2,2)\boldsymbol{R}_2^{M-k}, \quad 1 \le k \le M \end{aligned} \tag{17}$$

where

$$\begin{bmatrix} V(1,1) & V(1,2) \\ V(2,1) & V(2,2) \end{bmatrix} = -\begin{bmatrix} \boldsymbol{X} + \boldsymbol{R}_1\boldsymbol{B} & \boldsymbol{R}_1^{M-2}[\boldsymbol{F} + \boldsymbol{R}_1\boldsymbol{Y}] \\ \boldsymbol{R}_2^{M-2}[\boldsymbol{R}_2\boldsymbol{X} + \boldsymbol{B}] & \boldsymbol{R}_2\boldsymbol{F} + \boldsymbol{Y} \end{bmatrix}^{-1} \tag{18}$$

(ii) For $2 \le i \le M-2$, the i-th row is given by:

$$\boldsymbol{Z}(i,k) = \begin{cases} V(i,1)\boldsymbol{R}_1^{k-1} + V(i,2)\boldsymbol{R}_2^{i-k}, & 1 \le k \le i, \\ V(i,3)\boldsymbol{R}_1^{k-i-1} + V(i,4)\boldsymbol{R}_2^{M-k}, & i+1 \le k \le M, \end{cases} \tag{19}$$

and for $3 \le i \le M-1$, the i-th row is given by:

$$\boldsymbol{Z}(i,k) = \begin{cases} W(i,1)\boldsymbol{R}_1^{k-1} + W(i,2)\boldsymbol{R}_2^{i-k-1}, & 1 \le k \le i-1, \\ W(i,3)\boldsymbol{R}_1^{k-i} + W(i,4)\boldsymbol{R}_2^{M-k}, & i \le k \le M, \end{cases} \tag{20}$$

where

$$\begin{bmatrix} V(i,1) & V(i,2) & V(i,3) & V(i,4) \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \left(B^{(i)}[\boldsymbol{R}_1, \boldsymbol{R}_2]\right)^{-1}, \quad 2 \le i \le M-2$$
$$\begin{bmatrix} W(i,1) & W(i,2) & W(i,3) & W(i,4) \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \left(B^{(i-1)}[\boldsymbol{R}_1, \boldsymbol{R}_2]\right)^{-1}, \quad 3 \le i \le M-1$$

with for $2 \le i \le M-2$:

$$B^{(i)}[\boldsymbol{R}_1,\boldsymbol{R}_2] = \begin{bmatrix} \boldsymbol{X}+\boldsymbol{R}_1\boldsymbol{B} & -\boldsymbol{R}_1^i\boldsymbol{B} & \boldsymbol{R}_1^{i-1}\boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{R}_2^{i-2}\left[\boldsymbol{R}_2\boldsymbol{X}+\boldsymbol{B}\right] & \boldsymbol{R}_2\boldsymbol{F}+\boldsymbol{L} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{L}+\boldsymbol{R}_1\boldsymbol{B} & \boldsymbol{R}_1^{M-i-2}\left[\boldsymbol{F}+\boldsymbol{R}_1\boldsymbol{Y}\right] \\ \boldsymbol{0} & \boldsymbol{R}_2^{M-i-1}\boldsymbol{B} & -\boldsymbol{R}_2^{M-i}\boldsymbol{F} & \boldsymbol{R}_2\boldsymbol{F}+\boldsymbol{Y} \end{bmatrix}. \tag{21}$$

Observe that these direct formulas only hold for $M \ge 4$, however, if $M < 4$ it is beneficial to determine the inverse directly.

Acknowledgement

This research is supported by the Centre for Telematics and Information Technology (CTIT) of the University of Twente.

References

[1] N. Baer, R.J. Boucherie, and J.C.W. van Ommeren. Threshold queueing describes the fundamental diagram of uninterrupted traffic. Memorandum 2000, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, 2012.

[2] N. Baer, R.J. Boucherie, and J.C.W. van Ommeren. The $PH/PH/1$ multi-threshold queue. Memorandum 2011, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, 2013.

[3] W. Cao and W.J. Stewart. Iterative aggregation/dissagregation techniques for nearly uncoupled markov chains. *Journal of the Association for Computing Machinery*, 32(3):702–719, 1985.

[4] S.H. Choi, B. Kim, K. Sohraby, and B.D. Choi. On matrix-geometric solution of nested QBD chains. *Queueing Systems*, 43:5–28, 2003.

[5] B.N. Feinberg and S.S. Chiu. A method to calculate steady-state distributions of large markov chains by aggregating states. *Operations Research*, 35(2):282–290, 1987.

[6] D.P. Gaver, P.A. Jacobs, and G. Latouche. Finite birth-and-death models in randomly changing environments. *Advances in Applied Probability*, 16(4):715–731, 1984.

[7] M. Haviv. Aggregation/disaggregation methods for computing the stationary distribution of a markov chain. *SIAM Journal on Numerical Analysis*, 24(4):952–966, 1987.

[8] H.D. Kafeety, C.D. Meyer, and W.J. Stewart. A general framework for iterative aggregation/disaggregation methods. In *Proceedings of the Fourth Copper Mountain Conference on Iterative Methods*, 1992.

[9] M.N. Katehakis and L.C. Smit. A successive lumping procedure for a class of markov chains. *Probability in the Engineering and Informational Sciences*, 26(4):483–508, 2012.

[10] M.N. Katehakis, L.C. Smit, and F. Spieksma. Explicit solutions for a class of quasi-skip free in one direction processes. 2013.

[11] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer, 1960.

[12] D.S. Kim and R.L. Smith. An exact aggregation algorithm for a special class of markov chains. Technical report, University of Michigan, 1989.

[13] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability*, 30:650–674, 1993.

[14] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA., 1999.

[15] L.M. Le Ny and B. Tuffin. A simple analysis of heterogeneous multi-server threshold queues with hysteresis. In *Proceedings of the Applied Telecommunication Symposium*, 2002.

[16] S. Li and H. Sheng. Generalized folding algorithm for sojourn time analysis of finite qbd processes and its queueing applications. *Communications in Statistics - Stochastic Models*, 12(3):507–522, 1996.

[17] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*. Dover Publications, Inc., New York, 1981.

[18] D.F. Rogers and R.D. Plante. Estimating equilibrium probabilities for band diagonal markov chains using aggregation and disaggregation techniques. *Computers Operations Research*, 20(8):857–877, 1993.

[19] J. Ye and S. Li. Folding algorithm: A computational method for finite QBD processes with level-dependent transitions. *IEEE Trans. On Comm.*, 42(2):625–639, 1994.