Discrete time queueing networks with product form steady state. Availability and performance analysis in an integrated model

Christian Malchin, Hans Daduna

Queueing Syst (2010) 65: 385–421
DOI 10.1007/s11134-010-9181-2

Discrete time queueing networks with product form steady state. Availability and performance analysis in an integrated model

Christian Malchin · Hans Daduna

Received: 24 November 2006 / Revised: 17 May 2010 / Published online: 9 July 2010

Abstract For a discrete time network of generalized Bernoulli servers with unreliable nodes we derive the steady state probabilities for the joint queue length vector for all nodes and the availability status of the network. This allows us to assess the performance behavior and the reliability, resp. availability, of the network in an integrated model. Because our result exhibits a product form for the steady state distribution it opens the path to fast algorithmic evaluation of the desired performance and reliability indices.

Keywords Networks of discrete time queues · Unreliable nodes · Availability · Time reversed process · Product form theorem · Geometrical nodes

Mathematics Subject Classification (2000) 60K25 · 90B22 · 68M20

List of symbols

$\mathbb{N} = \{1, 2, \ldots\}$	Positive Integers.
$\mathbb{N}_0 = \{0, 1, 2, \ldots\}$.	
$C_j : \mathbb{N}_0 \to \mathbb{N}_0$	Capacity function; see pp. 6 and 24.
$S : M \to \mathbb{N}$	Route length; see p. 5.
$n(x), n({}^1x_j)$	Global and local queue length; see p. 8.
$A(x, (m, s)) =$	$\min\{t \mid s < t \leqq S(m), {}^2x_{w(m,t)} = 0\}$; jumping customer's new stage (upstream); see (10) on p. 18.
$\bar{A}(y, (m, s)) =$	$\max\{t \mid 1 \leqq t < s, {}^2y_{w(m,t)} = 0\}$; jumping customer's new stage (downstream); see (15) on p. 25.

This research is part of the DFG-research project "Stochastische Netzwerke in diskreter Zeit: Analyse von Leistung und Verfügbarkeit" (DA774/1-1).

C. Malchin · H. Daduna (✉)
University of Hamburg, Hamburg, Germany
e-mail: daduna@math.uni-hamburg.de

$\mathrm{UP}(x) = \{i \in \{1, 2, \ldots, J\} \mid {}^2x_i = 0\}$;
set of up-stations in state x; see (12) on p. 20.

$\mathrm{UP}_{\mathrm{OCC}}(x) = \{i \in \{1, 2, \ldots, J\} \mid n({}^1x_i) > 0, {}^2x_i = 0\}$,
set of active stations in state x; see (13) on p. 21.

$\mathrm{Succ}(x)$ Set of successor states of state x with respect to the original system (see p. 20).

$\overline{\mathrm{Succ}}(y)$ Set of successor states of state y with respect to the reversed system (see p. 28).

1 Introduction

In this note we study the asymptotic and stationary behavior of discrete time queueing networks. We extend product form results for discrete time open queueing networks [10, 12, 21, 43, 53], to include availability of unreliable nodes and state dependent arrival intensities.

In continuous time queueing network theory, product form calculus has been proven to be (possibly) the best route to efficient algorithms. Our contribution is in the same spirit: We provide explicit steady state distributions for the networks which (as in continuous time) show in equilibrium separability of the global state probabilities into factors which are determined by local characteristics only. Our main result shows, even more, that separability holds for the queue lengths as well as for the availability of the nodes in the networks.

The field of discrete time queueing theory has developed considerably since around fifteen years ago, motivated by the introduction of the ATM protocol in high-speed transmission networks; for a short introduction see [5, Chap. 4]. For a survey on discrete time networks see the books [5, 8, 10, 49]. The results described in these books are on classical topics of performance analysis for queueing systems and their networks.

Considering also availability of the nodes, with the model investigated in this note we bridge for discrete time networks the gap to the area of performability (= performance + reliability) where performance characteristics and availability (and their interaction) are studied in integrated models; for a survey see [18].

The model: We consider an open network with general topology. The nodes are generalized Bernoulli servers under First-Come-First-Served (FCFS). Customers arrive with an intensity which depends on the total population size in the network. They are of different types, and their type determines the route by which they want to pass through the network.

Nodes may break down randomly; repair time is random. Nodes which are up work properly; down nodes do not work at all, and customers are frozen there, while newly arriving customers bypass these nodes (this is the *jump over protocol* often introduced to resolve blocking in networks of queues; see [45]).

We further impose on the network model a regulation scheme that suppresses concurrent movements of customers. This regulation scheme is intimately connected to the so-called ALOHA protocol, [27, Sect. 5.11], [49, Sect. 6.2], which was introduced in transmission networks to resolve access conflicts to a shared medium by several users.

The main result of the paper is the steady state distribution of a Markov chain which describes jointly the state of all queues and the availability of all nodes. The steady state distribution is of product form and completely in line with the celebrated results for Jackson, Gordon–Newell, BCMP, and Kelly networks; see [9]. Therefore, similar to the product form theorems for pure queueing networks in continuous time and the subsequently derived computational algorithms, our result opens the path to jointly investigate the performance and availability of the network and their interdependence, and to subsequently develop computational algorithms for availability and performance.

Related work:

- Continuous time networks: The standard method to incorporate availability into product form models is the reduced service speed method; for a review of the classical method and refined versions see [7]. The resulting model is a pure queueing network where the unreliability of the servers is represented by slowing down the service rates. Integrated models for jointly investigating performance and availability in continuous time models are to be found, for example, in [39–41].
- Discrete time linear systems: Open tandems and closed cycles of completely reliable Bernoulli servers under FCFS are investigated in several papers. Some representatives are [23, 33, 34, 36]. A noteworthy observation in these linear networks is: We obtain product form steady state results without suppressing concurrent movements of customers.
- Unreliable stations: For the single unreliable Bernoulli server (often called unreliable Geo/Geo/1/∞ queue) and its refinements and generalizations there are many results available in the literature; see [5, Sect. 3.2] for a detailed exposition of a rather general one-node model with "server's interruption" due to breakdowns. More recent contributions in this direction are [14, 15, 47]. As expected, only for the simplest models explicit steady state distributions are obtained.

  In [29] availability was considered as an additional state characteristic in a linear system of unreliable Bernoulli servers and product form steady state results were obtained similar to those derived here.
- Discrete time analogs of Kelly networks of exponential and symmetric servers with general topology: These systems with completely reliable geometrical and doubly stochastic nodes date back to [12]; see [53] and [54] for more recent results similar to that. A survey is [10, Sect. 5].

  A distinguished property of these networks is (similar to those of Kelly- and the BCMP networks) that customers may be of different types and their path through the network can be traced for any of them individually, even on their dedicated places in the nodes and the servers individual customers are identifiable.

  In this general setting seemingly no explicit results for steady state distributions can be obtained when concurrent customer movements are possible, unless the network is linear.

  The geometrical servers introduced in [12] are the discrete time counterparts of Kelly's generalized exponential servers, [24, Sect. 3.1]. The present paper investigates networks of such geometrical servers which are unreliable.
- Concurrent movements of customers in discrete time networks: There is a rich class of models described in the literature, often under the heading of batch service,

batch arrivals, and batch movements, where customers indeed can move concurrently and some more or less explicit steady state distribution can be given; see [8] for a detailed review of discrete and continuous time systems. Explicit steady states are accessible here because it is not required that individual customers' passages through the network are followed, and special "departure rules" for groups of customers (of the same type, and for several types in parallel) are applied. Some representatives of such systems are dealt with in [2, 3, 19, 20, 31, 32], and from the related viewpoint of generalized semi-Markov processes in [21]. For a short review, emphasizing the necessity of investigating discrete time models, see the introductory section of [51].

In most of these papers continuous time is used and the authors propose that their methodology applies to discrete time systems as well: Simply by using the embedded jump chain as a model for its own. Applying this argument and procedure to Kelly networks with generalized exponential servers (as pointed out these are the counterpart of the geometrical servers) we obtain in a natural way discrete time networks without concurrent movements of customers.

A class of systems with batch service and batch movements, closely related to the models above, is the class of so-called *S-queues* introduced in [48].

Closely related to these models are the networks investigated in [50, 52], with applications to ATM networks.

- Discrete time stochastic automata networks: Such networks have many structural similarities with discrete time queueing networks. Their investigation suffers from the curse of combinatorial complexity in the same way as the networks considered here. In [16] it is shown that even for loosely coupled systems of automata one has to revert to approximations for computing the steady state distribution of the underlying Markov chains.
- Several further applications of models closely related to ours are described in some detail in Sect. 2.3.2 **(B)** below.

The "reversed process method" used to prove the main theorem:

Our method of proof is a standard device in queueing network theory since it was popularized in Kelly's book [24]—at least in continuous time. The method exploits the generator matrices of a stationary homogeneous Markov process (not necessarily reversible) and its time reversed process, which is a stationary homogeneous Markov chain again and has the same steady state; see [24, Sect. 1.7] for the general theory and with Theorem 1.13 (p. 30) as center of the method. We provide necessary details in Sect. 2.3.3 and quote the parallel theorem for discrete time.

In following the advice given by these theorems the main problem in proving our theorem was

- To construct the one-step transition matrix of the time reversed process (which can be interpreted as the state process of a rather weird network of queues), and
- To write down the matrices in a way that makes it possible to check the quasi-local balance equations, provided by the theorem, and
- In course of checking these equations to adapt product form formulas from the literature to the models investigated here

The rest of the paper is organized as follows: In Sect. 2 we describe the ingredients of our network and its construction in detail, state the main theorem, and discuss the

modeling assumptions and the methods applied in more detail. The formal description of the (as we shall call it: original) system is in Sect. 3; in Sect. 4 we describe formally the system which is the result of the time reversal of the original network process. The proof of the main theorem is postponed to Sect. 5.

2 System dynamics

2.1 Overview

We consider an open network of stations (queues) $Q[1], \ldots, Q[J]$, $J < \infty$, in discrete time $\mathbb{N}_0$.

Each station is a geometrical server (as introduced in [12]) with an infinite waiting room under First-Come-First-Served regime (FCFS).

Customers of different types arrive from the outside in state dependent arrival streams, travel through the network, and eventually depart to the outside. A customers' route is determined by his type according to Kelly's deterministic routing scheme [24, Sect. 3.1].

The servers are unreliable, break down randomly, and are repaired thereafter. Repair commences immediately after break down, the repair time is random. At stations with server under repair no service is provided and no further customers are admitted there. Customers who encounter on their path broken down servers will bypass these nodes.

2.2 The details

(1) Customer types, arrivals, and routing: The set M of customer types is countable; for simplicity we take either $M := \mathbb{N}$ or $M := \{1, \ldots, N\}$ for some $N \in \mathbb{N}$.

If at time $t \in \mathbb{N}_0$ the total population size in the network is $n \geqq 0$, then in the next time slot a customer of type $m \in M$ will arrive with probability $b(n) \cdot \alpha(m) \in (0, 1)$, where $\sum_{m \in M} \alpha(m) = 1$. With probability $c(n) = 1 - b(n) \in (0, 1)$ no customer will arrive. Given the total population size n the arrivals occur independently of the previous history of the network.

A customer of type m wants to pass the specified sequence of stations

$$Q\big[w(m, 1)\big], Q\big[w(m, 2)\big], \ldots, Q\big[w\big(m, S(m)\big)\big]$$

which are not necessarily distinct; in particular immediate feedback is possible. The *route* w is $w(m) := (w(m, 1), w(m, 2), \ldots, w(m, S(m)))$ and $S(m) \in \mathbb{N}$ is the *route length* for customers of type $m \in M$. Thus, type-m customers want to visit at *stage* s of their route station $Q[w(m, s)]$, $s = 1, \ldots, S(m)$.

Following Kelly [24] we call a pair (m, s) of a customer type $m \in M$ and a stage number $s \leqq S(m)$ a *class* of customer m.

(2) Geometrical servers: The stations (nodes) are geometrical servers with state dependent service capacity and ample waiting room. Such a station can be considered as a linearly ordered string of positions numbered $\{1, 2, 3, \ldots\}$. If n customers are present at station $Q[j]$, they reside in positions $\{1, 2, 3, \ldots, n\}$.

Then the *capacity function* $C_j : \mathbb{N}_0 \to \mathbb{N}_0$ determines the state dependent capacity of the station which is provided to the customers there:

The positions $\{1, \ldots, C_j(n)\}$ are busy, while positions $\{C_j(n) + 1, C_j(n) + 2, \ldots, n\}$ are idle. We require $1 \leqq C_j(n) \leqq n$ for $n \geqq 1$ and $C_j(0) = 0$. We consider C_j as a station parameter.

If n customers are present at time $t \in \mathbb{N}_0$ in station $Q[j]$, then for any customer on a busy position $1, \ldots, C_j(n)$ service ends (independently of another) with probability $p_j \in (0, 1]$ in the subsequent time slot, with probability $q_j = 1 - p_j$ this customer has to obtain service for at least one more time slot. The service times of all customers are mutually independent and are independent of the arrivals.

If n customers are present a new arrival occupies position $n + 1$ at the tail of the queue.

Reorganization of the queue is according to the shift protocol: If a customer's service expires and he departs from station $Q[j]$, jumping to the next station on his specified path, the gap in the line is closed, and customers behind him move up one position. In case of feedback the shift protocol applies before the customer is fed back.

There will be some additional restrictions imposed on the departure rules; see **(4)** below.

Examples Geometrical servers can be considered as discrete time analogs of Kelly's generalized exponential servers [24, Sect. 3.1]. We describe some examples:
(a) The standard multi-server with $s \geqq 1$ service channels is obtained by setting

$$C_1(n) = \begin{cases} n, & \text{for } 1 \leqq n \leqq s; \\ s, & \text{for } n \geqq s. \end{cases}$$

(b) A server with an additional service channel which is working only if the queue length n exceeds a prescribed critical value $d \geqq 1$ is obtained by setting

$$C_2(n) = \begin{cases} 1, & \text{for } 1 \leqq n \leqq d; \\ 2, & \text{for } n \geqq d + 1. \end{cases}$$

Two special cases of **(a)** should be mentioned:
(c) A single server is obtained by setting

$$C_3(n) = 1, \quad \text{for all } n \geqq 1.$$

(d) An infinite server is obtained by setting

$$C_4(n) = n, \quad \text{for all } n \geqq 1.$$

(3) Breakdown, repair, and availability: Breakdown and repair are determined locally at the individual nodes. Station $Q[j]$ is either *up* (0) or *down* (1). When a station is *up*, service can be provided to customers and customers can enter this station. When a station is *down*, no service is provided, old customers on their positions are frozen, no new customers are admitted, and the server is under repair.

The *availability* of station $Q[j]$, which indicates whether this station is *up* or *down*, is governed by a Markov chain with states $\{0, 1\}$ and one-step transition matrix $a_j = (a_j(k, \ell) : k, \ell \in \{0, 1\})$. These Markov chains act independently of one another and of the queue lengths of the nodes. Therefore servers may break down if they are idle, i.e., they are (in terms of reliability theory) in warm stand-by.

The network's availability is described by some vector $x \in \{0, 1\}^J$, where $x_j \in \{0, 1\}$ indicates whether node $Q[j]$ is *up* or *down*, $j \in \{1, \ldots, J\}$.

(4) Simultaneous departures and arrivals: Due to the discrete time scale, simultaneously several services may expire at one or more nodes and possibly an additional external arrival may occur at the same time instant. We impose the following rules in such a situation (for discussion of these rules, which will be termed **ALOHA protocol**, see Sect. 2.3.2):

- If there is only one event (departure or external arrival) at some time instant, it is executed immediately and scheduled according to the rules described above.
- If there are multiple events that occur simultaneously at the same time instant, none is granted: An arrival is rejected, any service that has expired is prolonged for at least one further time slot. The residual service of such blocked customers is provided according to the rules described in **(2)** independent of the system's history.

(5) Interaction of queueing processes and breakdown and repair: The simultaneous Markovian breakdown-repair processes develop free of restrictions, parallel (simultaneous) availability updates are feasible, and they occur in parallel to the state changes of the queueing processes.
Any change of the network's state is scheduled as follows:

- The queueing process develops one step taking into consideration the possible restrictions imposed by the nodes' availability: possible arrivals, departures, and inter-node transitions are performed under the side constraints prescribed by the ALOHA protocol **(4)**.
- Thereafter the network's availability is updated according to the Markovian rules described above: some down stations may be repaired, other down stations continue to be under repair, at other stations repair commences and customers are frozen there.

Summarizing: Reorganization of the queues depends on the current state only and is independent of the subsequent availability update.

(6) Rerouting in case of breakdowns: As described in **(3)** above, whenever a node is down and under repair, no new customers are admitted there. The consequence of this is that we have to implement a rerouting strategy for redirecting customers which on their route encounter a station which is under repair and therefore do not allow them to join the queue. We prescribe that a customer who is rejected by a station under repair skips this node without any time delay, and joins the next available station of his route. If there is none, he departs from the network. To be more specific:

If a customer of type m on his itinerary $w(m) = (w(m, 1), w(m, 2), \ldots, w(m, S(m)))$ finds station $Q[w(m, s)]$, $s < S(m)$, under repair, he immediately

tries to enter $Q[w(m,s+1)]$, and, if this station is under repair as well, and if $s+1<S(m)$, he immediately tries to enter $Q[w(m,s+2)]$, and so on, until he either finds a station $Q[w(m,t)]$, $s<t\leqq S(m)$, which is up, where he joins the queue, or, if there is no such station, he departs from the network immediately.

(7) State space description: The states of the network record for each station the sequence of customer classes (i.e. type and current stage) present there and the network's availability. We separate the queueing information ${}^1x$ and the availability information ${}^2x$ in the global state of the network

$$x=\left({}^1x,{}^2x\right)=\left(\left({}^1x_1,\dots,{}^1x_J\right),\left({}^2x_1,\dots,{}^2x_J\right)\right).$$

The second component ${}^2x=({}^2x_1,\dots,{}^2x_J)\in\{0,1\}^J$ of x describes the network's availability, where e.g., ${}^2x_j=0$ indicates that node $Q[j]$ is *up*.

The first component ${}^1x=({}^1x_1,\dots,{}^1x_J)$ of x encodes by ${}^1x_j$ for station $Q[j]$ for each position in the string of positions $1,2,\dots$ the class (m,s) of the customer who reside in this position.

A typical queueing state of the network therefore is ${}^1x=({}^1x_1,\dots,{}^1x_J)$, with

$$\begin{aligned}{}^1x_j&=\left({}^1x_{j,1},{}^1x_{j,2},\dots,{}^1x_{j,n({}^1x_j)},(0,0),(0,0),\dots\right)\\&=\left(\left({}^1x_{j,1,1},{}^1x_{j,1,2}\right),\left({}^1x_{j,2,1},{}^1x_{j,2,2}\right),\dots,\right.\\&\qquad\left.\left({}^1x_{j,n({}^1x_j),1},{}^1x_{j,n({}^1x_j),2}\right),(0,0),(0,0),\dots\right).\end{aligned}$$

The meaning of the components of ${}^1x_j$ is

- $n({}^1x_j)$ is the local queue length at $Q[j]$.
- $({}^1x_{j,1,1},{}^1x_{j,1,2})$ indicates that on the first position of station $Q[j]$ a type $m={}^1x_{j,1,1}$ customer resides, who is on stage $s={}^1x_{j,1,2}$ of his route through the network.
- $({}^1x_{j,n({}^1x_j),1},{}^1x_{j,n({}^1x_j),2})$ indicates that on the last occupied position of station $Q[j]$ a type $m={}^1x_{j,n({}^1x_j),1}$ customer resides, who is on stage $s={}^1x_{j,n({}^1x_j),2}$ of his route through the network.
- The intermediate positions can be described similarly.
- $(0,0)$ indicates that this position is not occupied. (Incorporating this redundant information for positions $n({}^1x_j)+1,n({}^1x_j)+2,\dots$ will simplify the presentation and several computations in the proofs, as will become visible later on: E.g., the queueing situation at station $Q[j]$ is described by a state in $(\mathbb{N}_0^2)^{\mathbb{N}}$.)

The set of all feasible queueing-availability states $x=({}^1x,{}^2x)$ of the network is $\hat{E}$, the set of all feasible queueing states ${}^1x$ is E. So $\hat{E}\subset E\times\{0,1\}^J$. We call ${}^1x_j$ the (local) queueing state at station $Q[j]$ and ${}^2x_j$ the (local) availability of station $Q[j]$. The total number of customers in system is $n(x):=\sum_{j=1}^J n({}^1x_j)$ (using $n(\cdot)$ for local and total queue length will not cause difficulties.)

An empty node has the queueing state $((0,0),(0,0),\dots)$. We set $n((0,0),(0,0),\dots)=0$.

Notational remark: For a concise notation of the concurrent state changes in the availability component (*availability update*) we use the following notation:

A change of the availability, say, at station $Q[j]$, can be described by the transition ${}^2x_j \to {}^2x_j \oplus 1$, where $\oplus$ denotes addition modulo 2 on $\{0,1\}$. To keep notation simple we denote the componentwise addition modulo 2 on $\{0,1\}^J$ by $\oplus$ as well.

We encode the network's availability update in a vector $u \in \{0,1\}^J$: if $u_j = 1$ the availability of $Q[j]$ will change, if $u_j = 0$ the availability remains unchanged. Then the network's availability ${}^2x$ and an availability update vector $u \in \{0,1\}^J$ result in the new network availability ${}^2y = {}^2x \oplus u$.

2.3 The queueing-availability process

From the description of the network in Sect. 2.2 it is easy to see that with state space $\hat{E}$ we can construct a Markov chain $X = (X_n : n \in \mathbb{N}_0)$ to describe the network's evolution over time.

With the help of the notation in **(7)** we can state our main result. (Recall that α is the type selection distribution for external arrivals; empty products are 1 by definition.)

Theorem 1 *Let X denote the homogeneous Markov chain for the queueing-availability process. If X is ergodic then the unique steady state and limiting distribution of X is given by*

$$\pi(x) = \frac{1}{K}\left(\prod_{k=1}^{n(x)} \frac{b(k-1)}{c(k)}\right)\prod_{l=1}^{J}\left(\prod_{k=1}^{n({}^1x_l)} \frac{\alpha({}^1x_{l,k,1})}{C_l(k)}\right)\frac{q_l^{n({}^1x_l)-C_l(n({}^1x_l))}}{p_l^{n({}^1x_l)}} a_l\big({}^2x_l \oplus 1, {}^2x_l\big),$$
$$x \in \hat{E}, \tag{1}$$

where K is the normalizing constant.

Remark The result of Theorem 1 is new even for the case of classical discrete time networks of completely reliable nodes; see [12, Theorem 1], [53], and [54], where the arrival streams to the network are assumed to be state independent, which simplifies analysis considerably.

2.3.1 Examples

We first consider single node systems with the service disciplines described in Sect. 2.2 **(2)**, and thereafter show how such nodes interact in networks.

(I) For simplicity of the presentation we always assume that there is only one customer type arriving, $M := \{1\}$, and that no feedback occurs, i.e. $S(1) = 1$, which leads to a simplified notation $E = \mathbb{N}_0$ for the set of all queueing states, and for the state space: $\hat{E} = \mathbb{N}_0 \times \{0,1\}$. We denote a typical state as $(n, {}^2x)$.

We further assume that the Bernoulli arrival stream is state independent, $b(n) = b$, $c(n) = c$ for all $n \in \mathbb{N}_0$.

Assuming ergodicity of the state process throughout we obtain the following steady states.

(a) For a multi-server with $s \geqq 1$ service channels

$$\pi\big((n,{}^2x)\big) = \frac{1}{K}\left(\frac{b}{c}\right)^n \frac{1}{\min(n,s)!}\left(\frac{1}{s}\right)^{(n-s)_+} \frac{q^{n-\min(n,s)}}{p^n} a\big({}^2x \oplus 1, {}^2x\big),$$
$$\big(n,{}^2x\big) \in \hat{E}, \tag{2}$$

(b) For a server with an additional service channel which is working only if the queue length n exceeds a prescribed critical value d

$$\pi\big((n,{}^2x)\big) = \frac{1}{K}\left(\frac{b}{c}\right)^n \left(\frac{1}{2}\right)^{(n-d)_+} \frac{q^{(n-1)_+}}{p^n}\left(\frac{1}{q}\right)^{1_{((n-d)_+>0)}} a\big({}^2x \oplus 1, {}^2x\big),$$
$$\big(n,{}^2x\big) \in \hat{E}, \tag{3}$$

(c) For a single server

$$\pi\big((n,{}^2x)\big) = \frac{1}{K}\left(\frac{b}{c}\right)^n \frac{q^{(n-1)_+}}{p^n} a\big({}^2x \oplus 1, {}^2x\big), \quad \big(n,{}^2x\big) \in \hat{E}, \tag{4}$$

(d) For an infinite server

$$\pi\big((n,{}^2x)\big) = \frac{1}{K}\left(\frac{b}{c}\right)^n \frac{1}{n!}\frac{1}{p^n} a\big({}^2x \oplus 1, {}^2x\big), \quad \big(n,{}^2x\big) \in \hat{E}. \tag{5}$$

Note, that in any case the steady state availabilities $a({}^2x \oplus 1, {}^2x)$ are obtained as the steady state of a two-state Markov chain, because we have $\{{}^2x \oplus 1, {}^2x\} = \{0, 1\}$ (set equality!).

(II) Our first network example is a linear three-stage tandem of single-server stations (see (4)). To obtain a simple example, we assume that there is only one type of arrivals and all customers intend to traverse the tandem stations subsequently via route $(1, 2, 3)$. The arrival stream is state-independent with arrival probability $b \in (0, 1)$, and the service probability at station $Q[j]$ is $p_j > b$, $j = 1, 2, 3$ (these relations guarantee ergodicity of the system).

Because of the simple routing and because there are no type distinctions we have a simplified notation for the set of all queueing states $E = \mathbb{N}_0^3$, and for the state space $\hat{E} = \mathbb{N}_0^3 \times \{0, 1\}^3$. We denote a typical state as $((n_1, n_2, n_3), ({}^2x_1, {}^2x_2, {}^2x_3))$.

The steady state according to Theorem 1 is, with some smoothing due to our simplifying assumptions and $a_l := a_l(0, 1) + a_l(1, 0)$,

$$\pi\big((n_1, n_2, n_3), \big({}^2x_1, {}^2x_2, {}^2x_3\big)\big)$$
$$= \prod_{l=1}^{3} a_l^{-1}\left(1 - \frac{b}{p_l}\right)\left(\frac{b}{1-b}\right)^{n_l} \frac{(1-p_l)^{(n_l-1)_+}}{p_l^{n_l}} a_l\big({}^2x_l \oplus 1, {}^2x_l\big),$$
$$\big((n_1, n_2, n_3), \big({}^2x_1, {}^2x_2, {}^2x_3\big)\big) \in \hat{E}. \tag{6}$$

Due to our simplifying assumptions in this example the normalizing constant factorizes as well and we therefore obtain independence of the queue lengths in steady state for a fixed time instant

$$K^{-1} = \prod_{l=1}^{3} K_l^{-1}, \quad K_l^{-1} = a_l^{-1}\left(1 - \frac{b}{p_l}\right).$$

Moreover, the normalizing constants factorize even for the queue lengths and the availabilities of the servers, a_l^{-1} is the normalization for the availability probabilities of station $Q[l]$, $l = 1, 2, 3$. The consequences of the observed factorization, resp., independence properties, are striking:

The joint steady state queue length distribution and the local steady state queue length distributions are independent of the breakdown probabilities and the speed of the repair at the stations and are the same as the station's steady state queue length distribution in isolation without breakdowns.

At node $Q[l]$ the steady state queue lengths probabilities are

$$\pi(n_l) = \left(1 - \frac{b}{p_l}\right)\left(\frac{b}{1-b}\right)^{n_l} \frac{(1-p_l)^{(n_l-1)_+}}{p_l^{n_l}}, \quad n_l \in \mathbb{N}_0.$$

The intuition with this observation is that during repair periods no arrivals are admitted at the node.

Nevertheless, there is degradation of the performance of the networks by unreliability, which is reflected, for example, in rewards earned by servicing customers. To be more specific: Let station $Q[l]$ obtain a reward r_l for servicing a customer successfully. Then the steady state reward per period (time unit) of station $Q[l]$ is

$$\begin{aligned} &\sum_{((n_1,n_2,n_3),({}^2x_1,{}^2x_2,{}^2x_3))\in\hat{E}} \pi\left((n_1,n_2,n_3),\left({}^2x_1,{}^2x_2,{}^2x_3\right)\right)\cdot 1_{(n_l>0,{}^2x_l=0)} r_l p_l \\ &= \frac{a_l(1,0)}{a_l}\cdot b\cdot r_l. \end{aligned} \tag{7}$$

If we set $r_l = 1$ we obtain from (7) the throughput at station $Q[l]$:

$$TH_l = \frac{a_l(1,0)}{a_l}\cdot b,$$

which is obviously strongly influenced by the mean repair time $a_l(1,0)^{-1}$.

Because in steady state the throughput equals the arrival rate it follows from Little's Theorem that individual customer's (mean) sojourn times are strongly affected by the availability and the repair time of the stations. The mean sojourn time $E(W_l)$ of a customer in steady state at station $Q[l]$, $l = 1, 2, 3$, is

$$E(W_l) = \frac{1-b}{p_l - b}\cdot\frac{a_l}{a_l(1,0)}.$$

We conclude that the mean sojourn times $E(W_l)$ of a customer and the local throughput TH_l at a node strongly depend on the local breakdown probability and on the

local mean repair times, but they do not depend on the breakdown probabilities and on the mean repair times at the other nodes. This is clearly a consequence of the simplifying assumptions which lead to an overall separability of the joint queue length—availability process.

(III) We now consider a network with three nodes, $Q[l]$, $l = 1, 2, 3$, and different customer types which is a generalization of the so-called Simon–Foley network.

Node $Q[1]$ is a multi-server with $s > 1$ service channels, node $Q[2]$ is a server with additional service channel which starts to work whenever the queue length exceeds the critical value d, and node $Q[3]$ is a single server.

The set of customer types is $M := \{1, 2, 3, 4\}$ and the routes are specified as follows:

- $w(1) := (1, 3)$.
- $w(2) := (1, 2)$.
- $w(3) := (2, 2, 3)$.
- $w(4) := (2)$.

So, customers of type $m \in \{1, 2, 3\}$ pass exactly two different nodes, but customers of type $m = 3$ have three stages because of (exactly) one feedback at node $Q[2]$, while customers of type $m \in \{1, 2\}$ both have two stages on their itinerary. Customers of type $m = 4$ visit only node $Q[2]$.

Assuming ergodicity, the steady state according to Theorem 1 then is

$$\begin{aligned}\pi(x) = {} & \frac{1}{K}\left(\prod_{k=1}^{n(x)} \frac{b(k-1)}{c(k)}\right)\left(\prod_{l=1}^{3}\left(\prod_{k=1}^{n({}^1x_l)} \alpha({}^1x_{l,k,1})\right) a_l\big({}^2x_l \oplus 1, {}^2x_l\big)\right) \\ & \times \frac{1}{\min(n({}^1x_1), s)!}\left(\frac{1}{s}\right)^{(n({}^1x_1)-s)_+} \frac{q^{n({}^1x_1)-\min(n({}^1x_1),s)}}{p^{n({}^1x_1)}} \\ & \times \left(\frac{1}{2}\right)^{(n({}^1x_2)-d)_+} \frac{q^{(n({}^1x_2)-1)_+}}{p^{n({}^1x_2)}}\left(\frac{1}{q}\right)^{1_{((n({}^1x_2)-d)_+>0)}} \frac{q^{(n({}^1x_3)-1)_+}}{p^{n({}^1x_3)}}, \qquad x \in \hat{E}.\end{aligned}$$

2.3.2 Discussion of the model and the result

(A) The regulation scheme for **simultaneous departures and arrivals** is in general a severe restriction imposed on the network's behavior, but seems to be common to almost all discrete time networks for which an explicit steady state distribution of product form is found; for a review see [10].

(1) We have in mind (at least) two general classes of networks where this regulation scheme is justified as modeling assumption:

- Low traffic networks where simultaneous events will be rare.
- Small time slots; then some additional slots of service will cause only small perturbations of global performance behavior.

We believe that in both cases performance predictions obtained by using the distribution described in the theorem will match well. This is in line with approximating

the Round–Robin queueing discipline, which is a regime for discrete time systems with multiple events, by the Processor Sharing discipline, a continuous time regime where multiple events in general do not occur. For discussions of this concept, see for example, [25, 37].

On the other hand, if these or similar conditions are not fulfilled, application of the models investigated here clearly is questionable without additional arguments.

(2) Further, the ALOHA protocol was developed for some by now classical applications, and recently got increasing interest in new applications. A typical example is as follows:

If several clients (transmission stations) share a common transmission medium and if concurrent usage of the medium perturbs all concurrent transmissions then none of these transmissions is successful.

A simple protocol which was implemented to deal with such transmission failures was the **slotted ALOHA protocol** which regulates the retrials of the clients by locally determined random schemes. See [27, Sect. 5.11], and [49, Sect. 6.2], for more details.

Our regulation scheme mimics exactly this protocol: successful transmission of customers (= messages) to the next station is possible if and only if exactly one service ends (= transmission request to some other node) or an external arrival (= new external transmission request) occurs exclusively.

We therefore refer to the regulation scheme **(4)** in Sect. 2.2 as **ALOHA protocol**.

(3) An important example, where steady states are of product form without this ALOHA-type restriction, are networks with linear structure: Open tandems and closed cycles of single server queues with state dependent service probabilities; see [10, Sects. 3 and 4].

Product form results for steady states of linear tandems with unreliable nodes can be found in [29].

A completely different approach to deal with simultaneous events is developed for Walrand's *S-queues* [48], where batch departures are compensated by batch arrivals without specifying positions in the queue.

Our experience with non-linear network topologies strongly suggests that there will be no product form result without additional modeling assumptions. This is supported by similar problems with a departure protocol used in [32] for regulating batch services. At every time epoch at most one node is selected to release a batch of customers to the system. As Miyazawa puts it, this protocol is motivated not only by the necessity to deal with discrete time networks, but also by its tractability for analysis.

In light of this discussion it is somewhat surprising that the main theorem allows for concurrent state changes of the availability vector without destroying product form steady state.

(B) The **schedule for rerouting customers that are blocked** because the destination node is broken down and does not admit new arrivals is exactly the so-called **jump over protocol**; see [45].

(1) The jump over protocol has become over the last decade a standard protocol to resolve blocking situations due to finite waiting rooms. An early paper using this

by-passing regime is [35], a more detailed discussion is in [42], where it is called **skipping**. Van Dijk [45] provides some intuitive arguments which support the conjecture that this protocol should lead to product form steady states for networks with blocking due to full waiting room.

More recent applications of this scheme can be found in [13] and [44, Chap. 3.6] (where it is called "blocking and rerouting").

Schassberger's results [42] on blocking networks with skipping were the starting point for an investigation of continuous time networks with unreliable nodes in [40] and [38]. The blocking scheme was transformed there into a scheme for handling unreliable nodes and their repair. This was introduced as *rerouting scheme in case of broken down nodes*. Other rerouting schemes for networks with unreliable stations can be found in [40] and [38] as well; these schemes will work in the present context too.

The introduction of the jump over protocol, resp. skipping protocol, was motivated mainly with the resulting product form steady state distributions for complex networks with blocking. It was often observed that in networks blocking of customers due to full stations leads to complicated steady state distributions, which are not accessible in explicit form. So simulation and numerical approximations have to be applied for performance analysis, if one insists on direct transformation of the real blocking protocols into the Markovian network models.

As an alternative approach, we can use approximating models, which lead to tractable analysis of a hopefully closely related model.

Another application of these approximating models is to obtain provable (exact) bounds for performance indices of networks. E.g., for mean values as throughput or mean time in system which are not accessible in the original networks we may find upper or lower bounds in comparing these quantities with those in related product form networks. A similar approach is possible for obtaining bounds of quantiles, utilizing stochastic orderings. An introduction to these principles is Chap. 4 of [46].

(2) Applications and related examples:
(i) *Loss networks*: Single and multi server loss systems have been investigated from the beginning of queueing theory, the most prominent formula is Erlang's loss formula $E_{1,m}(\lambda/\mu)$; see [26, p. 106]. This formula gives the loss probability at an $M/M/m/m$-system (m servers, no waiting room) for an arriving customer, who finds all servers occupied, i.e., the required resources are not available. These models and the results were generalized considerably to circuit-switching models. A survey of loss networks with product form steady state is [28]. The basic principle is described there as follows: "... customers arrive in attempt to seize some of the available system resources. At the time of arrival, a customer finding insufficient available resources leaves the system..." immediately, [28, p. 147].

Clearly, this principle applies in many situations, where insufficient available resources is a result of broken down servers. We sketch a simple example.

A network offers a variety of services, each service is represented by a service node. Customers indicate their requirement (assumed to be exactly one of the service types) by their type, so the route length is $S(m) = 1$ for all customers of "service" type m. Furthermore, these customers of type m enter then the node $Q[w(m, 1)]$ with $w(m, 1) := m$, which offers exactly this type of service.

Whenever on arrival of a type-m request node $Q[w(m,1)]$ is broken down, the arrival jumps over node $Q[w(m,1)]$ and leaves immediately the network, which is equivalent to being rejected. Assuming that the service is provided by geometrical servers, and that the arrivals occur in a (state dependent) Bernoulli process, this problem fits into the realm of Theorem 1. We clarify the situation with a network with the three nodes from our network example in Sect. 2.3.1; see Sect. 2.2 **(2)** for the definition of the service regimes.

Because the routing is now such that there is exactly one customer type at each node, $M := \{1,2,3\}$, and customers of type $m = j$ try to visit node $Q[j]$ only, it follows that we have a simplified notation for the set of all queueing states $E = \mathbb{N}_0^3$, and for the state space $\hat{E} = \mathbb{N}_0^3 \times \{0,1\}^3$. We denote a typical state as $((n_1,n_2,n_3),({}^2x_1,{}^2x_2,{}^2x_3))$.

Assuming ergodicity, the steady state according to Theorem 1 is (for $((n_1,n_2,n_3),({}^2x_1,{}^2x_2,{}^2x_3)) \in \hat{E}$)

$$\begin{aligned}
&\pi\big((n_1,n_2,n_3),\big({}^2x_1,{}^2x_2,{}^2x_3\big)\big)\\
&\quad= \frac{1}{K}\left(\prod_{k=1}^{n_1+n_2+n_3}\frac{b(k-1)}{c(k)}\right)\left(\prod_{l=1}^{3}\alpha(l)^{n_l}\right)\left(\prod_{l=1}^{3}a_l({}^2x_l\oplus 1,{}^2x_l)\right)\\
&\qquad\times\frac{1}{\min(n_1,s)!}\left(\frac{1}{s}\right)^{(n_1-s)_+}\frac{q^{n_1-\min(n_1,s)}}{p^{n_1}}\left(\frac{1}{2}\right)^{(n_2-d)_+}\frac{q^{(n_2-1)_+}}{p^{n_2}}\left(\frac{1}{q}\right)^{1_{((n_2-d)_+>0)}}\\
&\qquad\times\frac{q^{(n_3-1)_+}}{p^{n_3}}.
\end{aligned}\tag{8}$$

(ii) Facility reliability issues in network p-median problems were investigated by Berman, Krass, and Menezes [1]. They "... analyze a facility location model where facilities may be subject to disruptions, causing customers to seek service from operating facilities." They "... generalize the classical p-median problem on a network to explicitly include the failure probabilities, and analyze structural and algorithmic aspects of the resulting model." In the introduction of the paper they describe a series of real world scenarios where facility breakdowns happened due to various reasons—often causing an avalanche of breakdowns.

Their model is as follows: Customers arriving at a node of the network want to obtain service somewhere in the network; the main problem is where to position the (unreliable) servers (= facilities) and to associate customers to the facilities. The latter is done by specifying a (deterministic) sequence $(Q[w(m,1)], Q[w(m,2)], \ldots, Q[w(m,S(m))])$ (our notation) of different nodes where this customer asks for service, i.e., if facility $Q[w(m,1)]$ is disrupted (broken down), the customer jumps over to facility $Q[w(m,2)]$; if this is disrupted too, he jumps over to facility $Q[w(m,3)]$, etc.; when all facilities are disrupted the customer fails to be served.

There are costs specified with not servicing, and the aim is to minimize a generalized weighted distance for customers to arrive at a functioning facility.

The model is static in that no repair process is considered, and the service process and its duration are not taken into consideration. It should be noticed that nevertheless the resulting optimization problem is already extremely complex.

Our model opens a path to attack a dynamic version of this problem. We should note that in our present model a customer may obtain more than one services if there are more than one facility functioning on the path. Presumably, our model will therefore overestimate the total time in system for customers, and underestimate the throughput of the system. This is part of our ongoing research.

(iii) Flexible manufacturing systems (FMS) are from their very definition large and complex production systems with multiple plants and stores, where within each plant several production facilities (machines) are located, having a variety of specified product specific tools, and the additional property that machines are able to perform various production steps on various intermediate products. "Even more important, because of integrated computer control, jobs in the FMS can follow a rather flexible routing .... For instance, if one machine is not available (failed or occupied), the job can be routed to another machine to perform the same operation or another non-sequentially constrained operation" [6, p. 7]. So, machines can substitute other machines in the production process, if these are broken down or overloaded.

Our general model was developed in this direction, with the jump over protocol as a prescribed (rough and approximating) control policy to substitute broken down machines.

It should be noticed that modeling FMS by queueing networks usually describes the flexibility by random (Markovian) routing [6]. The transformation of our routing to this setting is not difficult because it is well known that random routing and deterministic routing are equivalent in that either routing scheme can be described (via type selection procedures) by using the other.

(iv) Further examples are discussed in [13, Sect. 2. Applications—examples] with respect to resolving blocking of customers.

2.3.3 Sketch of the proof

We shall utilize **the reversed process method**, which was popularized by Kelly in his book [24] to solve the complex steady state equations of continuous time systems, which arise from the interaction of many nodes of different structure and customers of different behavior. This method resembles the principle of local balance in time reversible processes. It exploits the fact that for stationary processes with state space E and transition intensities $q(x, y)$, $x, y \in E$, the time reversed process with state space E and transition intensities $\bar{q}(x, y)$, $x, y \in E$, has the same steady state distribution π as the original process. These transition intensities and the steady state probabilities are coupled by the quasi-local balance equations (note, that the process needs not to be reversible)

$$\pi(x)q(x, y) = \pi(y)\bar{q}(y, x) \quad \text{for all } x, y \in E.$$

The intensities $q(x, y)$ usually are given, the main problem is to guess how the time reversed process might look like: then write down the quasi-local balance equations. If we are able to solve with some $\bar{q}(\cdot, \cdot)$ (from our guess) these equations, we have found the steady state of both processes and in addition the intensities $\bar{q}(\cdot, \cdot)$ of the time reversed process as well. For more details see [24, Sect. 1.7].

The discrete time counterparts can be formulated in the following way [4, Theorem 6.1].

Let p be a stochastic matrix indexed by a countable set E, and let π be a probability distribution on E. Let $\bar{p}$ be any stochastic matrix indexed by E such that for all $i, j \in E$,

$$\pi(i)p(i,j) = \pi(j)\bar{p}(j,i). \tag{9}$$

Then π is a stationary distribution for p and $\bar{p}$. Moreover, if we consider a stationary homogeneous Markov chain X with one-step transition probability matrix p and initial distribution π, then $\bar{p}$ is the one-step transition probability matrix of the stationary homogeneous Markov chain which is obtained from X by reversing time.

Proof In a first step we describe in Sect. 3 in complete detail the transition probabilities of the network process. This is done in a three-step procedure:

1. We describe (deterministic) network transition operators which determine pathwise the physical transformations of the system.
2. We associate with these operators their respective occurrence probabilities when the state of the system is given; here we exploit explicitly the probabilistic assumptions put on the network.
3. We accumulate these probabilities to sum up to the one-step transition probabilities

$$p(x,y) \quad \text{for state transitions} \quad x \to y.$$

From our experience with previous work in [10, 29], and from similar proofs in the literature we have been able to guess

- The structure of the network under time reversal, and its transition probabilities, which will be described in a similar three-step procedure in Sect. 4.
- A candidate π for a product form steady state, which was announced in Theorem 1.

In fact, the expression for π, given in Theorem 1, was developed in parallel with the transition probabilities for the time reversed process in Sect. 4.2, following our experience with previous work in [10, 12, 29], and from similar proofs in the literature. □

The essential part of the proof in Sect. 5 relies on classifying pairs (x, y) according to the transition structure $x \to y$ within the network in Sect. 3.2 and $x \leftarrow y$ under time reversal in Sect. 4.2. We are able to partition suitably the squared state space $\hat{E} \times \hat{E}$ containing all such pairs (x, y). We show that within each partition subset the solution procedure of (9) follows the same lines for all (x, y) enclosed. This solution is presented in some detail, and we show that the procedure is in fact exhaustive.

3 Evolution of the network

We construct in this section the one-step transition probability matrix for the network process. This will be done with transition operators which describe the mechanics of

the network and thereafter combining these operators with the respective probabilities driving the network (for the continuous time analog see e.g. Kelly [24, Sect. 3.1]).

The transition operators are in fact only partial operators $T[\bullet]\colon D(T) \subseteq \hat{E} \to \hat{E}$ with domain $D(T)$. Whenever we write $T[\bullet](x)$ for some pair $x \in \hat{E}$, $T[\bullet]$, it is understood that $x \in D(T[\bullet])$ holds.

3.1 Network transitions

We need some notation for a concise description of the customers' movements, especially to describe jumps if stations are not accessible for customers.

Consider a customer of type $m \in M$ on stage $s \in \{1, \dots, S(m)\}$ of his path through the network, assume the network's state is $x \in \hat{E}$ and that this customer's service expires, and that he is allowed to depart from his present station $Q[w(m,s)]$ in accordance with the ALOHA protocol; see Sect. 2.2 **(4)**. If he finds station $Q[w(m,s+1)]$ in up status, i.e., ${}^2x_{w(m,s+1)} = 0$, he immediately enters that station. Otherwise according to the routing rules from Sect. 2.2 **(1)** he skips the subsequent broken down nodes on his path and jumps to station $Q[w(m,s')]$ where the customer's new stage s' is determined by

$$s' = A\big(x, (m,s)\big) := \min\big\{t \mid s < t \leqq S(m), {}^2x_{w(m,t)} = 0\big\}, \tag{10}$$

if such a station in up status exists, and otherwise he departs from the network. It will be convenient to define $A(x, (m,s)) := S(m) + 1$ in this case, and $w(m, S(m+1)) := J + 1$. Node $Q[J+1]$ is considered as the external sink: *entering node* $Q[J+1]$ means *leaving the network*.

We extend this notation for customer m arriving from the external source. Formally we set $w(m,0) = 0$ and consider $Q[0]$ as the external source with the associated capacity function $C_0 \equiv 1$.

We write $s' = A(x, (m,0))$ for the first stage on the path, such that the associated node $Q[w(m,s')]$ is up (here he joins the queue); if there is no such station, the arriving customer is rejected, and it will be convenient to write $A(x, (m,0)) := S(m) + 1$ again.

We call the direction of customers' movements with respect to increasing stage numbers *upstream*, and nodes before a customer on his residual path accordingly *upstream nodes*.

Customers' movements and availability update: In the following $x \in \hat{E}$ is a generic state of the network, and $u \in \{0,1\}^J$ is a vector that determines the network's availability update. Each state transition is governed by a compound transition operator which consists of successive applications of **(a)** a customers' movement operator, and **(b)** an availability update operator.

(a) Operators $T[\bullet]\colon D(T[\bullet]) \subset \hat{E} \to \hat{E}$ determine customers' movements (declared below by specifying $[\bullet]$). They leave the network's availability unchanged. So

$$x \mapsto T[\bullet](x)$$

results in

$${}^2\big(T[\bullet](x)\big) = {}^2x \quad \text{but usually} \quad {}^1\big(T[\bullet](x)\big) \neq {}^1x.$$

For simplicity we will erase the outer brackets

$$^2\big(T[\bullet](x)\big) =: {}^2T[\bullet](x) \quad \text{and} \quad {}^1\big(T[\bullet](x)\big) =: {}^1T[\bullet](x).$$

(b) The subsequent availability update is determined by $u \in \{0,1\}^J$.

With a little abuse of notation we denote by u as well the associated operator $u\colon \hat{E} \to \hat{E}$ with

$$x = \big({}^1x, {}^2x\big) \mapsto u(x) := \big({}^1x, {}^2x \oplus u\big).$$

Finally, the successor state of x is determined by applying the compound operator $u \circ T[\bullet]$, which results in:

$$x \mapsto \big(u \circ T[\bullet]\big)(x) = u\big(T[\bullet](x)\big) = \big({}^1T[\bullet](x), {}^2x \oplus u\big).$$

The following transitions of the network are feasible:

(i) **External arrival:** ARRIVAL OF A TYPE-m CUSTOMER FROM OUTSIDE AND AVAILABILITY UPDATE u.

We describe this transition by the compound operator $u \circ T[m]\colon \hat{E} \to \hat{E}$.

${}^1T[m](x) \in E$ is determined componentwise for $l \in \{1,\dots,J\}$.

If node $Q[l]$, $l = w(m, A(x,(m,0))) \in \{1,\dots,J\}$, is the first up node on m's path, then

$$\big({}^1T[m](x)\big)_l := \big({}^1x_{l,1}, {}^1x_{l,2}, \dots, {}^1x_{l,n({}^1x_l)}, \big(m, A\big(x,(m,0)\big)\big), (0,0), (0,0), \dots\big).$$

If $l \neq w(m, A(x,(m,0)))$, then

$$\big({}^1T[m](x)\big)_l := {}^1x_l.$$

Note that our definition of $A(x,(m,0))$ encompasses the case that the arriving customer is rejected, i.e. $w(m, A(x,(m,0))) = J+1$. This happens if ${}^2x_{w(m,s)} = 1$ for all $s \in \{1,\dots,S(m)\}$. No customer movement is visible in the state description: ${}^1T[m](x) = {}^1x$.

(ii) **Departure:** DEPARTURE OF THE CUSTOMER IN BUSY POSITION k OF STATION $Q[i]$ AND AVAILABILITY UPDATE u.

The network's transition is described by the compound operator $u \circ T[i,k]$.

The moving customer's destination node is $Q[w({}^1x_{i,k,1}, A(x,({}^1x_{i,k})))]$, and this determines which of the following cases occurs. The moving customer (a1) leaves the network, (a2) enters some other node, (b) re-enters the departure node (feedback).

The new queueing state ${}^1T[i,k](x)$ is determined componentwise for $l \in \{1,\dots,J\}$ as follows:

(a1) and (a2) Here $i \neq w({}^1x_{i,k,1}, A(x,({}^1x_{i,k}))) =: j$ (j may be the external sink). Then

$$\big({}^1T[i,k](x)\big)_l = \begin{cases} {}^1x_l & \text{if } l \neq i \text{ and } l \neq j, \\ ({}^1x_{i,1}, \dots, {}^1x_{i,k-1}, {}^1x_{i,k+1}, \dots, {}^1x_{i,n({}^1x_i)}, (0,0), \dots) \\ \quad \text{if } l = i. \end{cases}$$

And, if $Q[j]$ is not the sink,

$$\big({}^1T[i,k](x)\big)_j = \big({}^1x_{j,1},\ldots,{}^1x_{j,n({}^1x_j)},\big({}^1x_{i,k,1},A\big(x,\big({}^1x_{i,k}\big)\big)\big),(0,0),\ldots\big).$$

(b) Here $i = w({}^1x_{i,k,1}, A(x,({}^1x_{i,k})))$. (Note that possibly some intermediate nodes under repair on the customers' path have been short circuited.) Then for $l \neq i$

$$\big({}^1T[i,k](x)\big)_l := {}^1x_l,$$

and

$$\begin{aligned}\big({}^1T[i,k](x)\big)_i := \big({}^1x_{i,1},\ldots,{}^1x_{i,k-1},{}^1x_{i,k+1},\ldots,\\ {}^1x_{i,n({}^1x_i)},({}^1x_{i,k,1},A(x,({}^1x_{i,k}))),(0,0),\ldots\big).\end{aligned}$$

(iii) **No arrival, no departure:** AVAILABILITY UPDATE u, NO ARRIVAL, NO DEPARTURE.

This is described by the compound operator $u \circ T[\varnothing]$, where $T[\varnothing]$ is the identity operator on $\hat{E}$. So

$$\big(u \circ T[\varnothing]\big)(x) = u(x) = \big({}^1x, {}^2x \oplus u\big), \quad u \in \{0,1\}^J.$$

Remark: By denoting the identity operator on $\hat{E}$ as $T[\varnothing]$ we achieve a compact description in (11) below.

3.2 One-step transition probabilities

With the help of the partial compound operators $u \circ T[\bullet]$ we define the one-step transition probabilities

$$p = \big(p(x,y) : x,y \in \hat{E}\big).$$

We denote by $\mathrm{Succ}(x)$ the set of *successor states* of x, i.e. $y \in \mathrm{Succ}(x)$ if there exist some $u \circ T[\bullet]$ such that $(u \circ T[\bullet])(x) = y$.

A transition $x \to y \in \mathrm{Succ}(x)$ may be generated by several of the described operators. We denote by $\hat{p}(x, u \circ T[\bullet]) \geqq 0$ the probability that $x \to y$ is realized as $y = (u \circ T[\bullet])(x)$. It follows that

$$p(x,y) = \sum_{u \circ T[\bullet]} \hat{p}\big(x, u \circ T[\bullet]\big) \cdot \mathbf{1}\big[\big(u \circ T[\bullet]\big)(x) = y\big], \quad x,y \in \hat{E}, \tag{11}$$

where the summation runs over all $u \circ T[\bullet]$ such that $x \in D(T[\bullet])$.

For compact notation of the transition probabilities we need some abbreviations.

- We denote by $\mathrm{UP}(x)$ the set of up-stations in state x,

$$\mathrm{UP}(x) := \big\{i \in \{1,\ldots,J\} \mid {}^2x_i = 0\big\}. \tag{12}$$

- Further, $\mathrm{UP}_{\mathrm{OCC}}(x) \subset \mathrm{UP}(x)$ is the set of nodes that are active in state x, i.e. up and busy,

$$\mathrm{UP}_{\mathrm{OCC}}(x) := \bigl\{i \in \{1, \dots, J\} \mid {}^2x_i = 0,\ n\bigl({}^1x_i\bigr) > 0\bigr\}. \tag{13}$$

- If the network is in state x and a new customer of type m arrives from the external source, and if there is no node available (up) on his path then this customer is rejected. For state x we denote the set of these rejected customer types m as $\mathrm{REJ}(x)$.

The partial transition probabilities $\hat{p}(x, u \circ T[\bullet]) \geqq 0$ can now be classified according to Sect. 3.1 as follows. These probabilities can be written down directly following the detailed description in Sect. 2.2.

(i) **External arrival:** For $m \in M$ and $u \in \{0, 1\}^J$

$$\hat{p}\bigl(x, u \circ T[m]\bigr) := b\bigl(n(x)\bigr)\alpha(m)\Bigl(\prod_{l \in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\Bigr)\Bigl(\prod_{l=1}^{J} a_l\bigl({}^2x_l, {}^2x_l \oplus u_l\bigr)\Bigr);$$

(ii) **Departure:** For $i \in \mathrm{UP}_{\mathrm{OCC}}(x)$, $k \leqq C_i(n({}^1x_i))$ and $u \in \{0, 1\}^J$

$$\hat{p}\bigl(x, u \circ T[i, k]\bigr) := c\bigl(n(x)\bigr)\frac{p_i}{q_i}\Bigl(\prod_{l \in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\Bigr)\Bigl(\prod_{l=1}^{J} a_l\bigl({}^2x_l, {}^2x_l \oplus u_l\bigr)\Bigr);$$

(iii) **No arrival, no departure:** For $u \in \{0, 1\}^J$

$$\begin{aligned}
&\hat{p}\bigl(x, u \circ T[\varnothing]\bigr) \\
&\quad := \Bigl(1 - b\bigl(n(x)\bigr) \prod_{j \in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))} - c\bigl(n(x)\bigr) \\
&\qquad \times \sum_{i \in \mathrm{UP}_{\mathrm{OCC}}(x)} C_i\bigl(n\bigl({}^1x_i\bigr)\bigr)\frac{p_i}{q_i} \prod_{j \in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))}\Bigr) \\
&\qquad \times \Bigl(\prod_{l=1}^{J} a_l\bigl({}^2x_l, {}^2x_l \oplus u_l\bigr)\Bigr).
\end{aligned}$$

The one-step transition probabilities: We are now ready to accumulate all probabilities defined so far to obtain explicit expressions for the one-step transition probabilities of the network process. For a well-structured presentation we introduce a partition $\{\mathrm{O1}, \mathrm{O2}, \mathrm{O3a}, \mathrm{O3b}, \mathrm{O4}, \mathrm{O5}\}$ of the Cartesian state space product $\hat{E} \times \hat{E}$ and give closed form expressions for the one-step transition matrix p restricted to $\mathrm{O}i$, $i = 1, 2, 3\mathrm{a}, 3\mathrm{b}, 4, 5$. The partition is chosen in a way that construction of the $p(x, y)$ is uniform for all (x, y) in the same $\mathrm{O}i$, when i is fixed.

A first observation is that the availability update by operator u in a transition $x \to y$ is determined in any case by $u = {}^2x \oplus {}^2y$.

(O1) $(x, y) \in \mathrm{O1} :\iff y \in \mathrm{Succ}(x)$ and $n(y) - n(x) = 1$.
An external arrival has happened, because due to the ALOHA protocol $n(y) - n(x) = 1$ is equivalent to $n({}^1y_j) - n({}^1x_j) = 1$ for some $j \in \{1, \dots, J\}$ and $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1, \dots, J\} \setminus \{j\}$. Thus, an external customer of type $m = {}^1y_{j,n({}^1y_j),1}$ on position $n({}^1x_j) + 1$ at node $Q[j]$, with $j = w(m, A(x, (m, 0)))$ has arrived.

We therefore have $(u \circ T[m])(x) = y$ with $u = {}^2x \oplus {}^2y$, and this is the only possibility to generate transition $x \to y$. Hence

$$p(x, y) = b\big(n(x)\big)\alpha\big({}^1y_{j,n({}^1y_j),1}\big)\left(\prod_{l \in \mathrm{UP}(x)} q_l^{C_l(n(x_l))}\right)\left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)\right).$$

(O2) $(x, y) \in \mathrm{O2} :\iff y \in \mathrm{Succ}(x)$ and $n(y) - n(x) = -1$.
A departure to the exterior has happened, because due to the ALOHA protocol $n(y) - n(x) = -1$ is equivalent to $n({}^1y_i) - n({}^1x_i) = -1$ for some $i \in \{1, \dots, J\}$ and $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1, \dots, J\} \setminus \{i\}$. There is exactly one departure from $Q[i]$ to the outside. The class (m, s) of the departing customer can be determined by comparing x and y. In general, the departure position is not uniquely determined. We denote by

$$G_i^{\mathrm{O2}}\big({}^1x_i, {}^1y_i\big)$$

the set of possible departure positions.

Therefore, $(u \circ T[i, k])(x) = y$ holds if and only if $k \in G_i^{\mathrm{O2}}({}^1x_i, {}^1y_i)$. Hence

$$p(x, y) = c\big(n(x)\big)\big|G_i^{\mathrm{O2}}\big({}^1x_i, {}^1y_i\big)\big|\frac{p_i}{q_i}\left(\prod_{l \in \mathrm{UP}(x)} q_l^{C_l(n(x_l))}\right)\left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)\right).$$

Note that the departing customer at $Q[i]$ leaves the network if either $s = S(m)$ holds or all the upstream nodes (corresponding to stage numbers greater than s) on the route of the type m-customer are down.

(O3) $(x, y) \in \mathrm{O3} :\iff y \in \mathrm{Succ}(x)$ and $n(y) = n(x)$ and ${}^1x \neq {}^1y$.
An internal movement of some customer has happened; we separate O3 into two subsets.

(O3a) **No Feedback: Destination node $\neq$ departure node.** Due to the ALOHA protocol then $n(y) = n(x)$, and ${}^1x \neq {}^1y$ is equivalent to $n({}^1y_i) - n({}^1x_i) = -1$ for some $i \in \{1, \dots, J\}$, and $n({}^1y_j) - n({}^1x_j) = 1$ for some $j \in \{1, \dots, J\}$, $i \neq j$, and $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1, \dots, J\} \setminus \{i, j\}$.

Comparing x and y we determine:

- Departure node $Q[i]$ and destination node $Q[j]$.
- The set of possible departure positions $G_i^{\mathrm{O3a}}({}^1x_i, {}^1y_i)$. In general it is not possible to identify the departure position.

Now $(u \circ T[i,k])(x) = y$ holds, if and only if $k \in G_i^{\mathrm{O3a}}({}^1x_i, {}^1y_i)$. Therefore

$$p(x,y) = c\big(n(x)\big)\big|G_i^{\mathrm{O3a}}\big({}^1x_i, {}^1y_i\big)\big|\frac{p_i}{q_i}\left(\prod_{l\in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\right) \times \left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)\right). \tag{14}$$

(O3b) **Feedback.** A feedback has happened if and only if $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1,\dots,J\}$, and (due to the ALOHA protocol) ${}^1y_i \neq {}^1x_i$ for exactly one $i \in \{1,\dots,J\}$, and ${}^1y_j = {}^1x_j$ for all $j \in \{1,\dots,J\} \setminus \{i\}$.

Comparing x and y we determine:

- Feedback node $Q[i]$.
- The set of possible departure positions $G_i^{\mathrm{O3b}}({}^1x_i, {}^1y_i)$.

Now $(u \circ T[i,k])(x) = y$ holds, if and only if $k \in G_i^{\mathrm{O3b}}({}^1x_i, {}^1y_i)$. Hence

$$p(x,y) = c\big(n(x)\big)\big|G_i^{\mathrm{O3b}}\big({}^1x_i, {}^1y_i\big)\big|\frac{p_i}{q_i}\left(\prod_{l\in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\right) \times \left(\prod_{l=1}^{J} a_l({}^2x_l, {}^2y_l)\right).$$

(O4) $(x,y) \in \mathrm{O4} :\Longleftrightarrow y \in \mathrm{Succ}(x)$ and ${}^1y = {}^1x$.

This happens in a transition $x \to y$ if either no external arrival occurs and no service expires (where we have to incorporate the scheduling rules of the ALOHA protocol), or if an arriving external customer of type m is rejected because all nodes on his path are down: $m \in \mathrm{REJ}(x) \neq \varnothing$. Therefore

$$\begin{aligned}
&p(x,y)\\
&\quad= \Bigg(1 - b\big(n(x)\big)\prod_{j\in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))} - c\big(n(x)\big)\\
&\qquad\times \sum_{i\in \mathrm{UP_{OCC}}(x)} C_i\big(n\big({}^1x_i\big)\big)\frac{p_i}{q_i}\prod_{j\in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))}\\
&\qquad+ \sum_{m\in \mathrm{REJ}(x)} b\big(n(x)\big)\alpha(m)\prod_{l\in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\Bigg)\left(\prod_{l=1}^{J} a_l({}^2x_l, {}^2y_l)\right)\\
&\quad= \left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)\right)\Bigg(1 - b\big(n(x)\big)\sum_{m\in M\setminus \mathrm{REJ}(x)} \alpha(m)\prod_{j\in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))}\\
&\qquad- c\big(n(x)\big)\sum_{i\in \mathrm{UP_{OCC}}(x)} C_i\big(n\big({}^1x_i\big)\big)\frac{p_i}{q_i}\prod_{j\in \mathrm{UP}(x)} q_j^{C_j(n({}^1x_j))}\Bigg).
\end{aligned}$$

A little reflection shows that $p(x, y) = 0$ holds for all $(x, y) \notin (\mathrm{O1} \cup \mathrm{O2} \cup \mathrm{O3a} \cup \mathrm{O3b} \cup \mathrm{O4})$. We therefore define

$$\mathrm{O5} := \hat{E} \times \hat{E} \setminus (\mathrm{O1} \cup \mathrm{O2} \cup \mathrm{O3a} \cup \mathrm{O3b} \cup \mathrm{O4}),$$

and conclude that $\{\mathrm{O1}, \mathrm{O2}, \mathrm{O3a}, \mathrm{O3b}, \mathrm{O4}, \mathrm{O5}\}$ is clearly a partition of $\hat{E} \times \hat{E}$. Consequently, our construction of the one-step transition matrix is exhaustive.

4 Evolution of the time reversed network

As announced below, we want to apply time reversal for solving the steady state equations. So we have to guess the transition probabilities of the time reversed queueing-availability process and the steady state distribution. In continuous time network theory there often exists an appealing guess of a system (usually of similar characteristics) that is described by the time reversed process. Similar observation was made by the authors in some discrete time settings before. The present setting does not lead to such direct guess, as will be seen in the course of our description.

We call the network which is the result of our investigation here the (*time*) *reversed system* and distinguish from the network of Sects. 2 and 3 to which from now on we refer to as the *original system*. We shall see that the time reversed system can be identified considering three ingredients.

1. Time reversal of the isolated single station
2. Time reversal of the customers' routing, and
3. Time reversal of the interaction of routing, service, and availability

Our findings can be summarized as follows:

- The time reversed system is a queueing network with unreliable nodes, consisting of the nodes $Q[1], \ldots, Q[J]$.
- Customers arrive in a state dependent Bernoulli input stream with the same parameter as in the original system; types and type selection probabilities are as in the original system.

1. **The nodes.**
   - Every node has a waiting room of unlimited capacity, described by a sequence of positions $\{1, 2, 3, \ldots\}$.
   - At every node $Q[j]$ there is exactly one service position: departures occur from the occupied position with the highest number only.
     The probabilities for an ongoing service to expire are conditional Binomial distributions with suitably determined parameters $C_j(n)$ and p_j and will be described in detail below.
   - If there are n customers present at $Q[j]$ new arrivals are inserted into one of the entrance positions $\{1, 2, \ldots, C_j(n+1)\}$ at random.

   Comment: The multiple service positions in the nodes of the *original system* are mirrored by the multiple entrance positions in the nodes of the *reversed system*. Similarly, the single arrival position in the nodes of the *original system* is mirrored by the single departure position in the nodes of the *reversed system*.

2. **The routing.**
   - The routing for a customer of type m is to visit the same nodes as in the original system but in the reversed direction, i.e. with respect to **decreasing** stage numbers. We call the direction of the customers' movements with respect to decreasing stage numbers *downstream* and accordingly for a customer the residual nodes on his path the *downstream nodes*.
   - Broken down nodes are short circuited.

   Comment: Interchanging upstream and downstream progress of customers is intuitive and follows [24, Chap. 3]. Furthermore, the source $Q[0]$ and the sink $Q[J+1]$ can be thought to interchange their function.
3. **The interaction:**
   - The availability updates follow the same Markovian rules as in the original system (this is intuitive, because any two-state Markov chain is reversible).
   - Sequencing of movements and availability updating in a one-step transition are reversed: In a first step the availability of all nodes is updated, and thereafter the customers' movements are performed guided by the new availability status.

   Comment: Reversing the sequential actions clearly represents time reversal.

- Finally, we apply a scheduling of simultaneous events according to the ALOHA protocol.

4.1 Network transitions

Consider a customer of type $m \in M$ on stage $s \in \{S(m), S(m)-1, \dots, 1\}$ of his path through the network, assume the network's state is $y \in \hat{E}$ and that this customer's service expires, and that he is allowed to depart from his present station $Q[w(m,s)]$ in accordance with the rules given above. If he finds station $Q[w(m,s-1)]$ in up status, i.e., ${}^2y_{w(m,s-1)} = 0$, he immediately enters that station. Otherwise he skips the subsequent broken down nodes on his path and jumps to station $Q[w(m,s')]$ where the customer's new stage s' is determined by

$$\bar{A}\big(y,(m,s)\big) := \max\big\{t \mid 1 \leqq t < s, {}^2y_{w(m,t)} = 0\big\}, \tag{15}$$

if such a station in up status exists, and otherwise he departs from the network. We define $\bar{A}(y,(m,s)) := 0$ in this case. Recall that we have defined $w(m,0) = 0$ (see p. 402); now $Q[0]$ is considered as the external sink: entering node $Q[0]$ means *leaving the network*.

We extend this notation for customer m arriving from the external source (recall that we have defined $w(m,S(m)+1) = J+1$; see p. 402; we now consider $Q[J+1]$ as the external source). In this case $s' = \bar{A}(y,(m,S(m)+1))$ determines the first stage on the path, such that the associated node $Q[w(m,s')]$ is up (here he enters the queue); if there is no such station, the arriving customer is rejected; we write $\bar{A}(y,(m,S(m)+1)) := 0$ then. (Compare this with the definition of $A(x,(m,s))$ on p. 402: The attentive reader will notice that the roles of source and sink, i.e., $Q[0]$ and $Q[J+1]$, are interchanged, and the customer for determining his progress has to evaluate the "downstream nodes".)

Customer movements and availability update: In the following $y \in \hat{E}$ is a generic state of the network. Each state transition is governed by a compound transition operator which consists of successive applications of **(a)** firstly, an availability update operator, and **(b)** secondly, a customers' movements operator.

(a) The availability update is determined by some $u \in \{0,1\}^J$, and again with a little abuse of notation we denote by u as well the associated operator $u\colon \hat{E} \to \hat{E}$, which is exactly the same as that defined on p. 403.

(b) Operators $\bar{T}[\bullet]\colon D(\bar{T}[\bullet]) \subset \hat{E} \to \hat{E}$ determine customers' movements (declared below by specifying $[\bullet]$). They leave the network's availability unchanged.

$$y \mapsto \bar{T}[\bullet](y), \tag{16}$$

results in

$$^2\big(\bar{T}[\bullet](y)\big) = {}^2y \quad \text{but usually} \quad {}^1\big(\bar{T}[\bullet](y)\big) \neq {}^1y.$$

Similar to the case of the original system we erase the outer brackets:

$$^2\big(\bar{T}[\bullet](y)\big) =: {}^2\bar{T}[\bullet](y) \quad \text{and} \quad {}^1\big(\bar{T}[\bullet](y)\big) =: {}^1\bar{T}[\bullet](y).$$

Finally, the successor state of y is determined by applying the compound operator $\bar{T}[\bullet] \circ u$:

$$\begin{aligned} y \mapsto \big(\bar{T}[\bullet] \circ u\big)(y) &= \bar{T}[\bullet]\big(u(y)\big) = \big({}^1\bar{T}[\bullet]\big(u(y)\big), {}^2u(y)\big) \\ &= \big({}^1\bar{T}[\bullet]\big({}^1y, {}^2y \oplus u\big), {}^2y \oplus u\big). \end{aligned}$$

(Recall that in the original system we prescribed the sequencing $u \circ T[\bullet]$ for operating the network's transition.)

The following transitions of the network are feasible: For easier reading for a generic given y and u we will abbreviate $z := u(y)$. Note, that ${}^1z = {}^1y$ holds.

(i) **External arrival:** AVAILABILITY UPDATE u AND SUBSEQUENT ARRIVAL OF A TYPE-m CUSTOMER FROM OUTSIDE AT ENTRANCE POSITION k AT THE DESTINATION NODE.
We describe this transition by the compound operator

$$\bar{T}[m,k] \circ u. \tag{17}$$

Then the resulting network state is

$$\big(\bar{T}[m,k] \circ u\big)(y) = \big({}^1\bar{T}[m,k]\big(u(y)\big), {}^2u(y)\big). \tag{18}$$

Recall $z := u(y)$; we define ${}^1\bar{T}[m,k](z) \in E$ componentwise for $l \in \{1,\dots,J\}$; let $s' := \bar{A}(z,(m,S(m)+1)) \in \{1,\dots,S(m)\}$, $j := w(m,s')$ and $k \in \{1,2,\dots,C_j(n({}^1z_j)+1)\}$. Then

$$\big({}^1\bar{T}[m,k](z)\big)_l := {}^1z_l, \quad \text{if } l \neq j$$

$$\big({}^1\bar{T}[m,k](z)\big)_j := \big({}^1z_{j,1},\dots,{}^1z_{j,k-1},(m,s'),{}^1z_{j,k},\dots,{}^1z_{j,n({}^1z_j)},(0,0),\dots\big).$$

Whenever a customer is rejected (all nodes on his path are down), his destination is $Q[0]$; recall that we have defined $C_0 \equiv 1$, and therefore we have $k = 1$, and $\bar{T}[m, 1] \circ u$ is well defined in this case:

$$\bar{T}[m, 1] \circ u(y) = u(y).$$

(ii) **Departure:** AVAILABILITY UPDATE u, DEPARTURE FROM NODE $Q[i]$, AND SELECTION FOR ENTRANCE POSITION k AT THE DESTINATION NODE.

We describe this transition by the compound operator

$$\bar{T}[i; k] \circ u, \tag{19}$$

and distinguish three cases: The moving customer (a1) leaves the network, (a2) enters some other node, (b) re-enters the departure node (feedback). In any of these cases the moving customer's new stage is $s' = \bar{A}(u(y), {}^1y_{i,n({}^1y_i)})$ and his destination node therefore is $j = w(m, s')$ with $m = {}^1y_{i,n({}^1y_i),1}$.

(a1) The moving customer leaves the network. Recall $C_0 \equiv 1$, it follows for the coordinates of ${}^1\bar{T}[i; 1](z)$ that:

$$\left({}^1\bar{T}[i; 1](z)\right)_l := \begin{cases} {}^1z_l & \text{if } l \neq i, \\ ({}^1z_{i,1}, \ldots, {}^1z_{i,n({}^1z_i)-1}, (0, 0), \ldots) & \text{if } l = i. \end{cases}$$

(a2) Destination node $Q[j]$, $j \in \{1, \ldots, J\}$, is inside the network and no feedback occurs ($i \neq j$); the moving customer enters position k at $Q[j]$ (here $k \in \{1, 2, \ldots, C_j(n({}^1z_j) + 1)\}$ holds, because the capacity function C_j determines the number of possible entrance positions). The coordinates of ${}^1\bar{T}[i; k](z)$ are

$$\begin{aligned} \left({}^1\bar{T}[i; k](z)\right)_l &:= {}^1z_l, \quad \text{if } l \neq i \text{ and } l \neq j, \\ \left({}^1\bar{T}[i; k](z)\right)_i &:= \left({}^1z_{i,1}, \ldots, {}^1z_{i,n({}^1z_i)-1}, (0, 0), \ldots\right), \\ \left({}^1\bar{T}[i; k](z)\right)_j &:= \left({}^1z_{j,1}, \ldots, {}^1z_{j,k-1}, \left({}^1z_{i,n({}^1z_i),1}, s'\right), {}^1z_{j,k}, \ldots, \right. \\ &\qquad \left. {}^1z_{j,n({}^1z_j)}, (0, 0), \ldots\right). \end{aligned}$$

(b) A feedback occurs ($i = j$); the moving customer leaves node $Q[i]$, the gap is closed according to the shift protocol, and only thereafter he joins the queue again, occupying some entrance position $k \in \{1, 2, \ldots, C_i(n({}^1z_i))\}$ at $Q[i]$.

The coordinates of $\bar{T}[i; k](z)$ are

$$\begin{aligned} \left({}^1\bar{T}[i; k](z)\right)_l &:= {}^1z_l, \quad \text{if } l \neq i, \\ \left({}^1\bar{T}[i; k](z)\right)_i &:= \left({}^1z_{i,1}, \ldots, {}^1z_{i,k-1}, \left({}^1z_{i,n({}^1z_i),1}, s'\right), {}^1z_{i,k}, \ldots, \right. \\ &\qquad \left. {}^1z_{i,n({}^1z_i)-1}, (0, 0), \ldots\right). \end{aligned}$$

(iii) **No arrival, no departure:** AVAILABILITY UPDATE u. This is described by the compound operator $\bar{T}[\varnothing] \circ u$, where $\bar{T}[\varnothing]$ is the identity operator on $\hat{E}$. So

$$\big(\bar{T}[\varnothing] \circ u\big)(y) = u(y) = \big({}^1y, {}^2y \oplus u\big), \quad u \in \{0,1\}^J.$$

Remark: Here we denote the identity operator on $\hat{E}$ as $\bar{T}[\varnothing]$ to achieve a compact description in (20).

4.2 One-step transition probabilities

With the help of the partial compound operators $\bar{T}[\bullet] \circ u$ we define the one-step transition probabilities

$$\bar{p} = \big(\bar{p}(y,x) : x, y \in \hat{E}\big).$$

We denote by $\overline{\mathrm{Succ}}(y)$ the set of *successor states* of y, with respect to the time reversed system; i.e. $x \in \overline{\mathrm{Succ}}(y)$ if there exist some $\bar{T}[\bullet] \circ u$ such that $\bar{T}[\bullet] \circ u(y) = x$.

A transition $y \to x \in \overline{\mathrm{Succ}}(y)$ can be generated by several of the described operators: $x = \bar{T}[\bullet] \circ u(y)$. We denote by $\hat{\bar{p}}(y, \bar{T}[\bullet] \circ u) \geqq 0$ the probability that $y \to x$ is caused by an event that is described by $\bar{T}[\bullet] \circ u$.

The one-step transition matrix $\bar{p}$ on $\hat{E}$ is for $x, y \in \hat{E}$

$$\bar{p}(y,x) := \sum_{\bar{T}[\bullet] \circ u} \hat{\bar{p}}\big(y, \bar{T}[\bullet] \circ u\big) \cdot \mathbf{1}\big[\bar{T}[\bullet] \circ u(y) = x\big], \tag{20}$$

where the summation runs over all $\bar{T}[\bullet] \circ u$ such that $u(y) \in D(\bar{T}[\bullet])$.

The partial transition probabilities $\hat{\bar{p}}(y, \bar{T}[\bullet] \circ u)$ for the time reversed system can be classified as:

(i) **External arrival:** Let $Q[j]$ denote the arriving customers' destination node, where j is uniquely determined by y, u and m ($j = w(m, s')$ with $s' = \bar{A}(u(y), (m, S(m)+1))$). Then for $m \in M$, $u \in \{0,1\}^J$, and $k \in \{1, 2, \ldots, C_j(n({}^1y_j) + 1)\}$ (recall $C_0 \equiv 1$)

$$\hat{\bar{p}}\big(y, \bar{T}[m,k] \circ u\big) := \frac{b(n(y))\alpha(m)}{C_j(n({}^1y_j)+1)} \left(\prod_{l \in \mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))} \right) \left(\prod_{l=1}^{J} a_l({}^2y_l, {}^2y_l \oplus u_l) \right).$$

(ii) **Departure:** Let $Q[j]$ denote the jumping customers' destination node, which is uniquely determined by y and u. Then for $u \in \{0,1\}^J$, $i \in \mathrm{UP}_{\mathrm{OCC}}(u(y))$ and $k \in \{1, 2, \ldots, C_j({}^1n(y_j) + 1 - \delta(i,j))\}$

$$\hat{\bar{p}}\big(y, \bar{T}[i;k] \circ u\big) := \frac{c(n(y))C_i(n({}^1y_i))}{C_j(n({}^1y_j) + 1 - \delta(i,j))} \frac{p_i}{q_i} \left(\prod_{l \in \mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))} \right)$$
$$\times \left(\prod_{l=1}^{J} a_l({}^2y_l, {}^2y_l \oplus u_l) \right).$$

(iii) **No arrival, no departure:**

$$\hat{\bar{p}}\big(y, \bar{T}[\varnothing] \circ u\big) := \left(\prod_{l=1}^{J} a_l\big({}^2y_l, {}^2y_l \oplus u_l\big)\right) \cdot \left(1 - b\big(n(y)\big) \prod_{j \in \mathrm{UP}(u(y))} q_j^{C_j(n({}^1y_j))} - c\big(n(y)\big) \sum_{i \in \mathrm{UP}_{\mathrm{OCC}}(u(y))} C_i\big(n\big({}^1y_i\big)\big) \frac{p_i}{q_i} \prod_{j \in \mathrm{UP}(u(y))} q_j^{C_j(n({}^1y_j))}\right).$$

The one-step transition probabilities: We are now in a position to accumulate all probabilities obtained so far in a similar procedure as in the original system. We introduce similarly a partition $\{\mathrm{R1}, \mathrm{R2}, \mathrm{R3a}, \mathrm{R3b}, \mathrm{R4}, \mathrm{R5}\}$ of $\hat{E} \times \hat{E}$ and give closed form expressions for the one-step transition matrix $\bar{p}$ restricted to the Ri. Notice, "R" indicates time **R**eversed system, similar to the "O" types in the **O**riginal system. As the attentive reader will guess, we will eventually compare the transitions in an Oi with their counterparts in the Ri, and we will in any case show that such counterpart exists, and that the enumeration is exhaustive.

Similar to the original system, the availability update by operator u in a transition $y \to x$ is determined here in any case by $u = {}^2y \oplus {}^2x$.

(R1) $(y, x) \in \mathrm{R1} :\Longleftrightarrow x \in \overline{\mathrm{Succ}}(y)$ and $n(x) - n(y) = -1$.
A departure to the exterior has happened, because due to the ALOHA protocol $n(x) - n(y) = -1$ is equivalent to $n({}^1x_i) = n({}^1y_i) - 1$ for some $i \in \{1, 2, \ldots, J\}$ and $n({}^1x_j) = n({}^1y_j)$ for all $j \in \{1, 2, \ldots, J\} \setminus \{i\}$. Thus the departure occurred from node $Q[i]$. Only $\bar{T}[i; 1] \circ u$ triggers this transition, and therefore

$$\bar{p}(y, x) = c\big(n(y)\big) C_i\big(n\big({}^1y_i\big)\big) \frac{p_i}{q_i} \left(\prod_{l \in \mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))}\right) \left(\prod_{l=1}^{J} a_l({}^2y_l, {}^2x_l)\right).$$

(R2) $(y, x) \in \mathrm{R2} :\Longleftrightarrow x \in \overline{\mathrm{Succ}}(y)$ and $n(x) - n(y) = 1$.
An external arrival has happened, because due to the ALOHA protocol $n(x) - n(y) = 1$ is equivalent to $n({}^1x_j) = n({}^1y_j) + 1$ for some $j \in \{1, 2, \ldots, J\}$ and $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1, 2, \ldots, J\} \setminus \{j\}$. Thus, the arrival occurred at node $Q[j]$. The type m of the new customer can be determined by comparison of y and x. In general, the entrance position k can not be determined uniquely by comparison of y and x. We denote the set of possible entrance positions at $Q[j]$ under this transition by

$$\bar{G}_j^{\mathrm{R2}}\big({}^1y_j, {}^1x_j\big).$$

Exactly the operators $\bar{T}[m, k] \circ u$ with $k \in \bar{G}_j^{\mathrm{R2}}({}^1y_j, {}^1x_j)$ trigger this transition and therefore

$$\bar{p}(y, x) = b\big(n(y)\big) \alpha(m) \frac{|\bar{G}_j^{\mathrm{R2}}({}^1y_j, {}^1x_j)|}{C_j(n({}^1y_j) + 1)} \left(\prod_{l \in \mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))}\right) \times \left(\prod_{l=1}^{J} a_l({}^2y_l, {}^2x_l)\right).$$

(R3) $(y,x) \in \mathrm{R3} :\Longleftrightarrow x \in \overline{\mathrm{Succ}}(y)$ and $n(x) = n(y)$ and ${}^1x \neq {}^1y$.
Then an internal movement of some customer has happened; we split $R3$ into two subsets.

(R3a) **No Feedback: Destination node $\neq$ departure node.** In this case $n(x) = n(y)$ and ${}^1x \neq {}^1y$ is equivalent to $n({}^1x_j) = n({}^1y_j) - 1$ for some $j \in \{1,2,\dots,J\}$ and $n({}^1x_i) = n({}^1y_i) + 1$ for some $i \in \{1,2,\dots,J\}$, $i \neq j$, and $n({}^1y_l) = n({}^1x_l)$ for all $l \in \{1,2,\dots,J\} \setminus \{i,j\}$. This is due to a jump from node $Q[j]$ to node $Q[i]$. By comparing y and x, the departure node $Q[j]$ and the arrival node $Q[i]$ can be determined. The entrance position at $Q[i]$ can not be determined uniquely in general: we denote the set of possible entrance positions by $\bar{G}_i^{\mathrm{R3a}}({}^1y_i, {}^1x_i)$. Hence

$$\bar{p}(y,x) = c\big(n(y)\big)C_j\big(n\big({}^1y_j\big)\big)\frac{|\bar{G}_i^{\mathrm{R3a}}({}^1y_i, {}^1x_i)|}{C_i(n({}^1y_i)+1)}\frac{p_j}{q_j} \times \left(\prod_{l\in\mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))}\right)\left(\prod_{l=1}^{J} a_l\big({}^2y_l, {}^2x_l\big)\right). \tag{21}$$

(R3b) **Feedback.** In this case $n(x) = n(y)$ and ${}^1x \neq {}^1y$ is equivalent to $n({}^1x_l) = n({}^1y_l)$ for all $l \in \{1,2,\dots,J\}$ and ${}^1x_i \neq {}^1y_i$ for one and only one $i \in \{1,2,\dots,J\}$. Let $\bar{G}_i^{\mathrm{R3b}}({}^1y_i, {}^1x_i)$ be the set of possible entrance positions k. Now $x = \bar{T}[i;k] \circ u(y)$ holds if and only if $k \in \bar{G}_i^{\mathrm{R3b}}({}^1y_i, {}^1x_i)$. Consequently

$$\bar{p}(y,x) = c\big(n(y)\big)\big|\bar{G}_i^{\mathrm{R3b}}\big({}^1y_i, {}^1x_i\big)\big|\frac{p_i}{q_i}\left(\prod_{l\in\mathrm{UP}(u(y))} q_l^{C_l(n({}^1y_l))}\right) \times \left(\prod_{l=1}^{J} a_l\big({}^2y_l, {}^2x_l\big)\right).$$

(R4) $(y,x) \in \mathrm{R4} :\Longleftrightarrow x \in \overline{\mathrm{Succ}}(y)$ and ${}^1x = {}^1y$.
This happens within a transition $y \to x$ if either no external arrival occurs and no service expires (where we have to incorporate the scheduling rules of the ALOHA protocol) or if the arriving external customer of type m is rejected because after the availability update all nodes on his path are down. We denote the set of types m of these rejected customers by $\overline{\mathrm{REJ}}(u(y))$, where $u = y \oplus x$ is the availability update, associated with transition $y \to x$.
Now $\bar{T}[\varnothing] \circ u$ triggers the transition $y \to x$. But also the operators $\bar{T}[m,1] \circ u$, $m \in \overline{\mathrm{REJ}}(u(y))$, transform y into x. Hence

$$\bar{p}(y,x) = \left(\prod_{l=1}^{J} a_l\big({}^2y_l, {}^2x_l\big)\right) \times \left(1 - b\big(n(y)\big)\sum_{m\in M\setminus\overline{\mathrm{REJ}}(u(y))} \alpha(m) \prod_{j\in\mathrm{UP}(u(y))} q_j^{C_j(n({}^1y_j))} - c\big(n(y)\big)\sum_{i\in\mathrm{UP_{OCC}}(u(y))} C_i\big(n\big({}^1y_i\big)\big)\frac{p_i}{q_i}\prod_{j\in\mathrm{UP}(u(y))} q_j^{C_j(n({}^1y_j))}\right). \tag{22}$$

A little reflection shows that $\bar{p}(y,x) := 0$ holds for all $(y,x) \notin \mathrm{R1} \cup \mathrm{R2} \cup \mathrm{R3a} \cup \mathrm{R3b} \cup \mathrm{R4}$. We therefore define

$$\mathrm{R5} := \hat{E} \times \hat{E} \setminus (\mathrm{R1} \cup \mathrm{R2} \cup \mathrm{R3a} \cup \mathrm{R3b} \cup \mathrm{R4}),$$

and conclude that $\{\mathrm{R1}, \mathrm{R2}, \mathrm{R3a}, \mathrm{R3b}, \mathrm{R4}, \mathrm{R5}\}$ is clearly a partition of $\hat{E} \times \hat{E}$. Consequently, our construction of the one-step transition matrix is exhaustive for the time reversed process as well.

5 Proof of Theorem 1

Proof The reversed process method described in Sect. 2.3.3 requires to show

$$\pi(x)p(x,y) = \pi(y)\bar{p}(y,x) \quad \text{for all } x, y \in \hat{E}, \tag{23}$$

with (see Theorem 1)

$$\pi(x) = \frac{1}{K}\left(\prod_{k=1}^{n(x)} \frac{b(k-1)}{c(k)}\right)\prod_{l=1}^{J}\left(\prod_{k=1}^{n({}^1x_l)} \frac{\alpha({}^1x_{l,k,1})}{C_l(k)}\right)\frac{q_l^{n({}^1x_l)-C_l(n({}^1x_l))}}{p_l^{n({}^1x_l)}} a_l\big({}^2x_l \oplus 1, {}^2x_l\big). \tag{24}$$

When determining the transition matrices $(p(x,y) : x, y \in \hat{E})$ and $(\bar{p}(y,x) : x, y \in \hat{E})$ we classified the transition pairs $x \to y$ in the original system according to the partition $\{\mathrm{O1}, \mathrm{O2}, \mathrm{O3a}, \mathrm{O3b}, \mathrm{O4}, \mathrm{O5}\}$ of $\hat{E} \times \hat{E}$ and the transition pairs $x \leftarrow y$ in the time reversed system according to the partition $\{\mathrm{R1}, \mathrm{R2}, \mathrm{R3a}, \mathrm{R3b}, \mathrm{R4}, \mathrm{R5}\}$ of $\hat{E} \times \hat{E}$.

We shall utilize the fact that there is a natural bijection between members of the respective partition sets. In fact the partition elements are connected by the relation

$$(x,y) \in \mathrm{O}i \iff (y,x) \in \mathrm{R}i, \quad i = 1, 2, 3\mathrm{a}, 3\mathrm{b}, 4, \tag{25}$$

(hence, $(x,y) \in \mathrm{O5} \iff (y,x) \in \mathrm{R5}$).

Then we prove for $i = 1, 2, 3\mathrm{a}, 3\mathrm{b}, 4, 5$

$$\pi(x)p(x,y)|_{\mathrm{O}i} = \pi(y)\bar{p}(y,x)|_{\mathrm{R}i} \quad \text{for all } (x,y) \in \mathrm{O}i \quad (\text{and therefore } (y,x) \in \mathrm{R}i)$$

(the restrictions $p|_{\mathrm{O}i}$ and $\bar{p}|_{\mathrm{R}i}$ of the one-step transition matrix are given in Sects. 3.2 and 4.2).

We explain in full detail the proof for the case where $x \to y$ is due to a customer's transition from a network node to some other network node, i.e., $(x,y) \in \mathrm{O3a}$, and refer to [30] for the other cases.

Let $(x,y) \in \mathrm{O3a}$ and let the transition $x \to y$ (in the original system) be triggered by a customer of class (m,s) departing from node $Q[i]$ and jumping to node $Q[j]$, where $i \neq j$, and let k be this customer's departure position at $Q[i]$. It follows $(m,s) = {}^1x_{i,k}$, $i = w(m,s)$, and $j = w(m,s')$ with $s' = A(x,(m,s))$.

Recall (p. 406) that by definition $k \in G_i^{\mathrm{O3a}}({}^1x_i, {}^1y_i)$, the set of possible departure positions.

Recall further, that for the transition $x \to y$ there exists a uniquely determined $u \in \{0,1\}^J$ such that $u \oplus {}^2x = {}^2y$ and $u \oplus {}^2y = {}^2x$.

From this description we can immediately deduce that $(y,x) \in$ R3a:

Because of $n({}^1y_j) \geqq 1$ and ${}^2(u(y)) = u \oplus {}^2y = {}^2x$ (and therefore ${}^2(u(y))_j = 0$) we can apply $\bar{T}[j;k]$ to transform $u(y)$ as follows.

On position $n({}^1y_j)$ at node $Q[j]$ resides a class (m,s') customer, $s' = A(x,(m,s))$. His stage after the time reversed jump (according to $\bar{T}[j;k] \circ u(y)$) is $s'' = \bar{A}(u(y),(m,s')) = \max\{t \mid 0 \leqq t < s', {}^2x_{w(m,t)} = 0\}$ and obviously $s'' = s$ holds. This implies that the destination node of the jumping customer (in reversed time) is indeed $Q[i]$. There the jumping customer is inserted into position k, yielding $x = (\bar{T}[j;k] \circ u)(y)$ with $u = {}^2x \oplus {}^2y$.

We notice *en passant* that exactly those positions $k \in \{1,\dots,C_i({}^1x_i)\}$, where a class (m,s) customer can depart from $Q[i]$ to join the tail of the class sequence ${}^1y_j$ at $Q[j]$ to perform the transformation $x \to y$ in the original system, are the positions where in the time reversed system the class (m,s) customer can be inserted at $Q[i]$ who departed as the class (m,s') customer from the single service position at $Q[j]$ to perform the transition $x \leftarrow y$. This yields

$$G_i^{\mathrm{O3a}}\big({}^1x_i, {}^1y_i\big) = \bar{G}_i^{\mathrm{R3a}}\big({}^1y_i, {}^1x_i\big).$$

Inserting (24) and (14), the left side of (23) is (we abbreviate in the formulas below $\underline{J} := \{1,2,\dots,J\}$)

$$\begin{aligned}
&\pi(x)p(x,y)\\
&\quad= \frac{1}{K}\left(\prod_{k=1}^{n(x)} \frac{b(k-1)}{c(k)}\right)\prod_{l=1}^{J}\left(\prod_{k=1}^{n({}^1x_l)} \frac{\alpha({}^1x_{l,k,1})}{C_l(k)}\right)\frac{q_l^{n({}^1x_l)-C_l(n({}^1x_l))}}{p_l^{n({}^1x_l)}} a_l\big({}^2x_l \oplus 1, {}^2x_l\big)\\
&\qquad\cdot c\big(n(x)\big)\big|G_i^{\mathrm{O3a}}\big({}^1x_i, {}^1y_i\big)\big|\frac{p_i}{q_i}\left(\prod_{l\in \mathrm{UP}(x)} q_l^{C_l(n({}^1x_l))}\right)\left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)\right)\\
&\quad= \frac{1}{K}\left(c\big(n(x)\big)\prod_{k=1}^{n(x)} \frac{b(k-1)}{c(k)}\right)\left(\prod_{l=1}^{J} a_l\big({}^2x_l, {}^2y_l\big)a_l\big({}^2x_l \oplus 1, {}^2x_l\big)\right)\\
&\qquad\times\left(\big|G_i^{\mathrm{O3a}}\big({}^1x_i, {}^1y_i\big)\big|\left(\frac{q_i}{p_i}\right)^{n({}^1x_i)-1}\prod_{k=1}^{n({}^1x_i)} \frac{\alpha({}^1x_{i,k,1})}{C_i(k)}\right)\\
&\qquad\times\left(\left(\frac{q_j}{p_j}\right)^{n({}^1x_j)}\prod_{k=1}^{n({}^1x_j)} \frac{\alpha({}^1x_{j,k,1})}{C_j(k)}\right)\\
&\qquad\times\left(\prod_{l\in \mathrm{UP}(x)\setminus\{i,j\}}\left(\frac{q_l}{p_l}\right)^{n({}^1x_l)}\prod_{k=1}^{n({}^1x_l)} \frac{\alpha({}^1x_{l,k,1})}{C_l(k)}\right)\\
&\qquad\times\left(\prod_{l\in \underline{J}\setminus \mathrm{UP}(x)} \frac{q_l^{n({}^1x_l)-C_l(n({}^1x_l))}}{p_l^{n({}^1x_l)}}\prod_{k=1}^{n({}^1x_l)} \frac{\alpha({}^1x_{l,k,1})}{C_l(k)}\right).
\end{aligned} \tag{26}$$

Inserting (24) and (21), the right side of (23) is

$$
\begin{aligned}
&\pi(y)\bar{p}(y,x)\\
&\quad=\frac{1}{K}\left(\prod_{k=1}^{n(y)}\frac{b(k-1)}{c(k)}\right)\prod_{l=1}^{J}\left(\prod_{k=1}^{n({}^1y_l)}\frac{\alpha({}^1y_{l,k,1})}{C_l(k)}\right)\frac{q_l^{n({}^1y_l)-C_l(n({}^1y_l))}}{p_l^{n({}^1y_l)}}a_l\big({}^2y_l\oplus 1,{}^2y_l\big),\\
&\qquad\times c\big(n(y)\big)C_j\big(n\big({}^1y_j\big)\big)\frac{|\bar{G}_i^{\mathrm{R3a}}({}^1y_i,{}^1x_i)|}{C_i(n({}^1y_i)+1)}\frac{p_j}{q_j}\\
&\qquad\times\left(\prod_{l\in\mathrm{UP}(u(y))}q_l^{C_l(n({}^1y_l))}\right)\left(\prod_{l=1}^{J}a_l({}^2y_l,{}^2x_l)\right)\\
&\quad=\frac{1}{K}\left(c\big(n(y)\big)\prod_{k=1}^{n(y)}\frac{b(k-1)}{c(k)}\right)\left(\prod_{l=1}^{J}a_l\big({}^2y_l,{}^2x_l\big)a_l\big({}^2y_l\oplus 1,{}^2y_l\big)\right)\\
&\qquad\times\left(C_j\big(n\big({}^1y_j\big)\big)\left(\frac{q_j}{p_j}\right)^{n({}^1y_j)-1}\prod_{k=1}^{n({}^1y_j)}\frac{\alpha({}^1y_{j,k,1})}{C_j(k)}\right)\\
&\qquad\times\left(\frac{|\bar{G}_i^{\mathrm{R3a}}({}^1y_i,{}^1x_i)|}{C_i(n({}^1y_i)+1)}\left(\frac{q_i}{p_i}\right)^{n({}^1y_i)}\prod_{k=1}^{n({}^1y_i)}\frac{\alpha({}^1y_{i,k,1})}{C_i(k)}\right)\\
&\qquad\times\left(\prod_{l\in\mathrm{UP}(u(y))\setminus\{i,j\}}\left(\frac{q_l}{p_l}\right)^{n({}^1y_l)}\prod_{k=1}^{n({}^1y_l)}\frac{\alpha({}^1y_{l,k,1})}{C_l(k)}\right)\\
&\qquad\times\left(\prod_{l\in\underline{J}\setminus\mathrm{UP}(u(y))}\frac{q_l^{n({}^1y_l)-C_l(n({}^1y_l))}}{p_l^{n({}^1y_l)}}\prod_{k=1}^{n({}^1y_l)}\frac{\alpha({}^1y_{l,k,1})}{C_l(k)}\right).
\end{aligned}
\tag{27}
$$

We check the equality of (26) and (27) by comparing line by line.

- **First line:** The products in the first huge brackets are equal because the total population size is not changed in course of the transformation $x\to y$, so $n(x)=n(y)$. Equality of the products in the second huge brackets can be seen by formal manipulation, because for all $u,v\in\{0,1\}$ we have

$$a_l(u,v)a_l(u\oplus 1,u)=a_l(v,u)a_l(v\oplus 1,v)$$

  which for $u\neq v$ follows from $a_l(0,1)a_l(0\oplus 1,0)=a_l(0,1)a_l(1,0)=a_l(1,0)\times a_l(1\oplus 1,1)$.
- **Fourth and fifth line:** The transition in the reversed process $x\leftarrow y$ starts with updating the availability and yields ${}^2u(y)={}^2x$, especially $\mathrm{UP}(x)=\mathrm{UP}(u(y))$; furthermore, at nodes other than $Q[i]$, $Q[j]$ the local queueing state is not changed, i.e., ${}^1x_l={}^1y_l$ for $l\in\underline{J}\setminus\{i,j\}$. Combining these observations, we can directly check equality of the third lines.

- **Second and third line:** The specification of the jump of a class-(m,s)-customer from $Q[i]$ as a class-(m,s')-customer to $Q[j]$ and vice-versa yields $n({}^1y_j)-1=n({}^1x_j)$ and $n({}^1y_i)+1=n({}^1x_i)$. Also, the jumping customer's *type* remains unchanged, therefore it follows that

$$\alpha\big({}^1y_{j,n({}^1y_j),1}\big)\prod_{k=1}^{n({}^1y_i)}\alpha\big({}^1y_{i,k,1}\big)=\prod_{k=1}^{n({}^1x_i)}\alpha\big({}^1x_{i,k,1}\big),$$

because the customers on positions $1,2,\ldots,n({}^1y_i)$ at node $Q[i]$ are not involved in the transition, and ${}^1y_{j,n({}^1y_j),1}=m$ is the jumping customer's type. Furthermore, we already noticed that

$$G_i^{\mathrm{O3a}}\big({}^1x_i,{}^1y_i\big)=\bar{G}_i^{\mathrm{R3a}}\big({}^1y_i,{}^1x_i\big).$$

Combining these observations we transform the second line of (27):

$$\begin{aligned}
&\left(C_j\big(n({}^1y_j)\big)\left(\frac{q_j}{p_j}\right)^{n({}^1y_j)-1}\prod_{k=1}^{n({}^1y_j)}\frac{\alpha({}^1y_{j,k,1})}{C_j(k)}\right)\\
&\quad\times\left(\frac{|\bar{G}_i^{\mathrm{R3a}}({}^1y_i,{}^1x_i)|}{C_i(n({}^1y_i)+1)}\left(\frac{q_i}{p_i}\right)^{n({}^1y_i)}\prod_{k=1}^{n({}^1y_i)}\frac{\alpha({}^1y_{i,k,1})}{C_i(k)}\right)\\
&=\left(\left(\frac{q_j}{p_j}\right)^{n({}^1x_j)}\left(\prod_{k=1}^{n({}^1x_j)}\frac{\alpha({}^1y_{j,k,1})}{C_j(k)}\right)\right)\\
&\quad\times\left(\big|\bar{G}_i^{\mathrm{R3a}}\big({}^1y_i,{}^1x_i\big)\big|\left(\frac{q_i}{p_i}\right)^{n({}^1x_i)-1}\frac{\alpha({}^1y_{j,n({}^1y_j),1})\prod_{k=1}^{n({}^1y_i)}\alpha({}^1y_{i,k,1})}{\prod_{k=1}^{n({}^1x_i)}C_i(k)}\right)\\
&=\left(\big|G_i^{\mathrm{O3a}}\big({}^1x_i,{}^1y_i\big)\big|\left(\frac{q_i}{p_i}\right)^{n({}^1x_i)-1}\prod_{k=1}^{n({}^1x_i)}\frac{\alpha({}^1x_{i,k,1})}{C_i(k)}\right)\\
&\quad\times\left(\left(\frac{q_j}{p_j}\right)^{n({}^1x_j)}\prod_{k=1}^{n({}^1x_j)}\frac{\alpha({}^1x_{j,k,1})}{C_j(k)}\right),
\end{aligned}$$

which is the second line of (26). □

6 Conclusion

We have investigated in this paper a discrete time open network of queues with unreliable nodes, which in our opinion is an important model because already its *reliable* version and variants of it have found applications, e.g., in telecommunication and production systems. Our aim was to find an analog to the celebrated product form steady state distributions which are well known in the equilibrium theory of queueing networks in continuous time as well as in discrete time.

The positive message obviously is that even in discrete time networks with unreliable nodes product form modeling is possible when breakdown and repair are incorporated into an integrated model. On the other hand it turned out that the proofs are extremely tedious and generalizations towards more complicated networks seem to be not easy.

Nevertheless, there is still much research to be done. We hope that our ongoing research in this area will contribute to further modeling techniques to obtain product form network scenarios for networks with unreliable servers and different rerouting schemes.

A counterpart of our main theorem holds for closed networks of queues with unreliable nodes. The proof is by rewriting the proof of the present paper along the lines of the proofs which are given in Sect. 5.6 in [10] for the case of reliable nodes.

Part of our present research is concerned with networks with discrete time symmetric servers; see [11, 12, 43, 53, 54] for the case of reliable network nodes.

An interesting question arises from the observation that our main work in the proof was to establish the structure of the time reversed process and the associated network: Is it possible to develop a discrete time analog of the continuous time *Reversed Compound Agent Theorem* (RCAT) of Harrison (for a short description and generalizations see [17]), or of the *compositional approach to performance modeling* of Hillston [22], which would then produce to a certain extent automatically the time reversed model which we obtained in the present paper by the reversed process method directly?

We believe that the answer is in the affirmative (not proposing to see the final construction), but this would need much more work. The main point will be: Complexity of simultaneous-event description, and handling concurrent state changes at different network nodes. So the question is still open whether this would reduce the amount of notation and the handling of combinatorial complexity of the system presented here.

Acknowledgements We thank three anonymous referees for their careful reading of the manuscript and their constructive critics which (as we hope) made the paper much more easily accessible to the audience.

Especially, the possible connection of our method of proof to the (continuous time) *Reversed Compound Agent Theorem* (RCAT) of Harrison and the *compositional approach to performance modeling* of Hillstonism was remarked by one of the referees.

References

1. Berman, O., Krass, D., Menezes, M.B.C.: Facility reliability issues in network p-median problems: Strategic centralization and co-location effects. Oper. Res. **55**, 332–350 (2007)
2. Boucherie, R.J., van Dijk, N.M.: Spatial birth–death processes with multiple changes and applications to batch service networks and clustering processes. Adv. Appl. Probab. **22**, 433–455 (1990)
3. Boucherie, R.J., van Dijk, N.M.: Product forms for queueing networks with state-dependent multiple job transitions. Adv. Appl. Probab. **23**, 152–187 (1991)
4. Bremaud, P.: Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Texts in Applied Mathematics, vol. 31. Springer, New York (1999)
5. Bruneel, H., Kim, B.G.: Discrete-Time Models for Communication Systems Including ATM. Kluwer Academic, Boston (1993)
6. Buzacott, J.A., Yao, D.D.: On queueing network models of flexible manufacturing systems. Queueing Syst. Appl. **1**, 5–27 (1986)

7. Chakka, R., Mitrani, I.: Approximate solutions for open networks with breakdowns and repairs. In: Kelly, F.P., Zachary, S., Ziedins, I. (eds.) Stochastic Networks, Theory and Applications. Royal Statistical Society Lecture Notes Series, vol. 4, pp. 267–280. Clarendon, Oxford (1996). Chap. 16
8. Chao, X., Miyazawa, M., Pinedo, M.: Queueing Networks—Customers, Signals, and Product Form Solutions. Wiley, Chichester (1999)
9. Chen, H., Yao, D.D.: Fundamentals of Queueing Networks. Springer, Berlin (2001)
10. Daduna, H.: Queueing Networks with Discrete Time Scale: Explicit Expressions for the Steady State Behavior of Discrete Time Stochastic Networks. Lecture Notes in Computer Science, vol. 2046. Springer, Berlin (2001)
11. Daduna, H., Schassberger, R.: A discrete-time round-robin queue with Bernoulli input and general arithmetic service time distributions. Acta Inform. **15**, 251–263 (1981)
12. Daduna, H., Schassberger, R.: Networks of queues in discrete time. Z. Oper. Res. **27**, 159–175 (1983)
13. Economou, A., Fakinos, D.: Product form stationary distributions for queueing networks with blocking and rerouting. Queueing Syst. Appl. **30**, 251–260 (1998)
14. Fiems, D., Steyaert, B., Bruneel, H.: Analysis of a discrete-time GI-G-1 queueing model subjected to bursty interruptions. Comput. Oper. Res. **30**(1), 139–153 (2003)
15. Fiems, D., Steyaert, B., Bruneel, H.: Discrete-time queues with generally distributed service times and renewal-type server interruptions. Perform. Eval. **55**(3–4), 277–298 (2004)
16. Gusak, O., Dayar, T., Fourneau, J.-M.: Stochastic automata networks and near complete decomposability. SIAM J. Matrix Anal. Appl. **23**(2), 581–599 (2001)
17. Harrison, T.T., Lee, P.G.: Separable equilibrium state probabilities via time reversal in Markovian process algebra. Theor. Comput. Sci. **346**(1), 161–182 (2005)
18. Haverkort, B.R., Marie, R., Rubino, G., Trivedi, K.: Performability Modeling, Technique and Tools. Wiley, New York (2001)
19. Henderson, W., Pearce, C.E.M., Taylor, P.G., van Dijk, N.M.: Closed queueing networks with batch services. Queueing Syst. Appl. **6**, 59–70 (1990)
20. Henderson, W., Northcote, B.S., Taylor, P.G.: Triggered batch movement in queueing networks. Queueing Syst. Appl. **21**, 125–141 (1995)
21. Henderson, W., Pearce, C.E.M., Taylor, P.G., van Dijk, N.M.: Insensitivity in discrete-time generalized semi-Markov processes allowing multiple events and probabilistic service scheduling. Ann. Appl. Probab. **5**, 78–96 (1995)
22. Hillston, J.: A compositional approach to performance modelling. PhD thesis, University of Edinburgh (1994)
23. Hsu, J., Burke, P.J.: Behaviour of tandem buffers with geometric input and markovian output. IEEE Trans. Commun. **24**, 358–361 (1976)
24. Kelly, F.P.: Reversibility and Stochastic Networks. Wiley, Chichester (1979)
25. Kleinrock, L.: Time-shared systems: A theoretical treatment. J. Assoc. Comput. Mach. **14**(2), 242–261 (1967)
26. Kleinrock, L.: Queueing Theory, vol. I. Wiley, New York (1975)
27. Kleinrock, L.: Queueing Theory, vol. II. Wiley, New York (1976)
28. Kobayashi, H., Mark, B.L.: Product-form loss networks. In: Dshalalow, J.H. (ed.) Frontiers in Queueing, Models and Applications in Science and Engineering. Probability and Stochastics Series, pp. 147–195. CRC Press, Boca Raton (1997). Chap. 6
29. Malchin, C., Daduna, H.: Availability and performance analysis in a discrete time tandem network with product form steady state. In: German, A., Heindl, R. (eds.) Proceedings of the GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communications Systems, pp. 381–398. VDE-Verlag, Berlin (2006)
30. Malchin, C., Daduna, H.: Discrete time queueing network with product form steady state: Availability and performance analysis in an integrated model. Preprint 2006-02, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, Fachbereich Mathematik der Universität Hamburg (2006, submitted), 40 p.
31. Miyazawa, M.: On the characterisation of departure rules for discrete–time queueing networks with batch movements and its applications. Queueing Syst. Appl. **18**, 149–166 (1994)
32. Miyazawa, M.: Stability of discrete-time Jackson networks with batch movements. In: Glasserman, P., Sigman, K., Yao, D.D. (eds.) Stochastic Networks: Stability and Rare Events. Lecture Notes in Statistics, vol. 117, pp. 76–93. Springer, New York (1996). Chap. 5
33. Pestien, V., Ramakrishnan, S.: Asymptotic behavior of large discrete–time cyclic queueing networks. Ann. Appl. Probab. **4**, 591–606 (1994)

34. Pestien, V., Ramakrishnan, S.: Features of some discrete-time cyclic queueing networks. Queueing Syst. Appl. **18**, 117–132 (1994)
35. Pittel, B.: Closed exponential networks of queues with saturation: the Jackson-type stationary distribution and its asymptotic analysis. Math. Oper. Res. **4**, 357–378 (1979)
36. Pujolle, G., Claude, J.P., Seret, D.: A discrete queueing system with a product form solution. In: Hasegawa, T., Takagi, H., Takahashi, Y. (eds.) Proceedings of the IFIP WG 7.3 International Seminar on Computer Networking and Performance Evaluation, pp. 139–147. Elsevier, Amsterdam (1986)
37. Sakata, M., Noguchi, S., Oizumi, J.: An analysis of the M/G/1 queue under round-robin scheduling. Oper. Res. **19**, 371–385 (1971)
38. Sauer, C.: Stochastic product form networks with unreliable nodes: Analysis of performance and availability. PhD thesis, University of Hamburg, Department of Mathematics (2006)
39. Sauer, C., Daduna, H.: Separable networks with unreliable servers. In: Charzinski, J., Lehnert, R., Tran-Gia, P. (eds.) Providing QoS in Heterogeneous Environments. Teletraffic Science and Engineering, vol. 5b, pp. 821–830. Elsevier, Amsterdam (2003)
40. Sauer, C., Daduna, H.: Availability formulas and performance measures for separable degradable networks. Econ. Qual. Control **18**, 165–194 (2003)
41. Sauer, C., Daduna, H.: Degradable networks with general up and down time distributions. In: Buchholz, P., Lehnert, R., Pioro, M. (eds.) Proceedings of the 12th GI/ITG Conference on MMB Together with 3rd PGTS, pp. 185–194. VDE-Verlag, Berlin (2004)
42. Schassberger, R.: A new approach to the M/G/1 processor-sharing queue. Adv. Appl. Probab. **16**, 202–213 (1984)
43. Schassberger, R., Daduna, H.: A discrete–time technique for solving closed queueing network models of computer systems. In: Kühn, P.J., Schulz, K.M. (eds.) Messung, Modellierung und Bewertung von Rechensystemen. Informatik–Fachberichte, vol. 61, pp. 122–134. Springer, Berlin (1983)
44. Serfozo, R.F.: Introduction to Stochastic Networks. Applications of Mathematics, vol. 44. Springer, New York (1999)
45. van Dijk, N.M.: On Jackson's product form with 'jump-over' blocking. Oper. Res. Lett. **7**(5), 233–235 (1988)
46. van Dijk, N.M.: Queueing Networks and Product Forms—a Systems Approach. Wiley, Chichester (1993)
47. Vinck, B., Bruneel, H.: System delay versus system content for discrete-time queueing systems subject to server interruptions. Eur. J. Oper. Res. **175**, 362–375 (2006)
48. Walrand, J.: A discrete–time queueing network. J. Appl. Probab. **20**, 903–909 (1983)
49. Woodward, M.E.: Communication and Computer Networks: Modelling with Discrete-Time Queues. IEEE Comput. Soc., Los Alamitos (1994)
50. Woodward, M.E.: Product-form closed discrete-time queueing networks with finite capacity shared buffer nodes. Electron. Lett. **32**(20), 1875–1876 (1996)
51. Woodward, M.E.: Size-limited batch movement in product-form closed discrete-time queueing networks. ACM SIGMTRICS Perform. Eval. Rev. **25**(1), 139–146 (1997)
52. Woodward, M.E.: Product form solutions for discrete-time queueing networks with bursty traffic. Electron. Lett. **36**(17), 1512–1514 (2000)
53. Yates, R.D.: High speed round-robin queueing networks. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (1990)
54. Yates, R.D.: Analysis of discrete time queues via the reversed process. Queueing Syst. Appl. **18**, 107–116 (1994)