U N I V E R S I T Ä T H A M B U R G

BCMP networks with unreliable servers

C. Sauer, H. Daduna

Preprint–No. 2003-01 February 2003

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

BCMP networks with unreliable servers

Cornelia Sauer and Hans Daduna

University of Hamburg
Department of Mathematics
Center of Mathematical Statistics
and Stochastic Processes

February 2003

Abstract: We model and analyse stochastic networks with unreliable servers. Such models have various applications e.g. in telecommunication networks and production systems.
Starting from standard Jackson, Gordon-Newell and BCMP networks, we introduce degradable network models which incorporate the breakdowns of the servers and their repairs into the Markovian system description. The breakdown event may occur in different ways: Nodes may break down as an isolated event or in groups simultaneously, and the repair of the nodes may end for each node individually or in groups as well. Working and repair periods are assumed to be governed by state dependent intensities.
We apply three different rules to handle the customers' routing connected with nodes in repair status. These rules originate from communication and / or production theory where they are used to resolve blocking situations.
Our main results: Modeling unreliable networks in this way leads to steady state distributions of product form for the supplemented Markovian state processes.

MSC (2000) subject classification: 60 K 25, 90 B 25, 60 K 37
Key words: Performability, reliability, Jackson networks, Gordon-Newell networks, BCMP networks, blocking principles, product form solution

1 Introduction

It is well known that single and multichannel servers with unreliable service channels pose hard problems when explicit and simply to use formulas are required. There are early papers in the area of Operations Research which contribute to this field, see e.g. [WC58], [Kei62], [Thi63], [AN63] for the single server case and [MAI68] and [MW94] for the case of a node with parallel servers. In [Mik88] and [MIKS95] a two-node-network with unreliable nodes is analysed and it is shown, that considerable mathematical effort using the theory of complex variables is needed. A systematic study of single node systems with unreliable service channels is [BFH88].

On the other hand there is a queueing networks success story which began with the work of Jackson, Gordon and Newell, the BCMP–group, and Kelly, who provided us up to 1975 with a versatile class of so-called *separable* networks with explicitly accessible steady state of *product form.* For recent surveys see [Ser99], [BGdT98], [Dad01].

But it turned out that occurrence of unreliable transmission channels (or nodes) in networks destroys the elegance of the performance analysis methods based on product form solutions. Network models with nodes of degradable performance which we found in the literature are not separable. In [MCl88] an approximate product form expression of the steady state for a class of separable queueing networks, in which the servers can operate on different levels of degradation, is derived using the observed property of near-complete decomposability and the consequences thereof.

A recent survey on how to approximately solve models which introduce the feature of unreliable nodes into queueing network models is [CM96], where four different suggestions to overcome the lack of explicit steady state solutions are recollected and developed further.
• For the reduced work–rate approximation the unreliable server is replaced by a reliable one with the same service capacity which incorporates the mean down times.
• The Poisson approximation replaces the interacting network nodes by nodes in isolation with suitably adjusted Poisson arrivals and then uses the single server results from classical theory.
• The MMPP approximation (Markov–modulated Poisson process [FMH93]) generalizes the latter approach by integrating the non–renewal property of the network flows into the arrival processes of the isolated queues.
• A further modification is the joint–state approximation, where the up/down status of the nodes is controlled by an environment process and then incorporated into the MMPP approximation.

Our approach is different from those described in [CM96]. We develop new network models which integrate breakdowns of the servers and their repair into the very first Markovian system description. The probability for a breakdown of a server particularly depends on the locally defined load of that node, and the repair process may be state dependent as well. We obtain for these new network models steady states of product form.

Our work can be classified to lay in the field of PERFORMABILITY THEORY which unifies performance analysis and reliability theory, see [Mey84] or [HMRT01]. The recent proceedings [GLT01] contains theoretical studies as well as applications of performability techniques mainly in the fields of computer systems and telecommunications networks. The interplay between performance analysis and reliability theory is well known to be a matter of structural complexity and computational problems. Indeed both stochastic processes, namely that describing the performance of the system and that describing the availability of the system's components, are processes in a random environment, defined by the respective other process. These interacting processes, resp. random environments, are usually thought to resist explicit computations without strong simplifications (see e.g. the reduced work-rate approximation). During the last two decades we have observed that as a response to this performability theory was invented as a result of the need to model and evaluate computer and communication networks and distributed manufacturing systems which exhibit degradable performance. Therefore the aim of our research is to further contribute to bridge the gap between the worlds of PERFORMANCE ANALYSIS and RELIABILITY THEORY in connection with queueing networks.

We introduce in Section 2 as our starting point the classical exponential and BCMP networks with stochastic routing and nonexponential service times. In Section 3 we describe the breakdown and repair mechanisms and introduce the candidates for the steady state distributions for the degradable networks in Section 4. Sections 5 and 6 constitute the paper's core: After introducing the used routing principles in case of a server's breakdown, we state and proof our main results on product form steady states. In Section 7 we derive in special cases availability and performance measures for the degradable BCMP networks. Section 8 contains some conclusions and an outlook.

For simplicity of the presentation and for a better understanding of the underlying ideas throughout the paper we first develop our results in easy-to-handle exponential networks and thereafter extend them to the much more general BCMP networks which require some more notational burdening. The main difficulty is to develop modified routing regimes which regulate the routing when there are nodes in repair status by forbidding new customers to enter these nodes.

Further specifications of the breakdown and repair intensities and further routing protocols in case of breakdowns which yield (modified) product form steady states for networks with unreliable servers are subject of ongoing research.

Throughout the paper we set empty sums to 0, empty products to 1, and $\frac{0}{0} := 0$.

2 Standard separable networks

In this section we summarize the required definitions and theorems which will be needed in the sequel. The starting point for general queueing network theory were the open Jackson networks of single server nodes (with state dependent service intensities), which were introduced by Jackson [Jac57]. We use these networks to demonstrate the fundamental principles of our approach to performability. The relevant definition is as follows.

Definition 2.1 (Jackson network) *A Jackson network is a network of service stations (nodes) numbered $\{1, 2, \ldots, J\} := \bar{J}$. Station j is a single server with infinite waiting room under FCFS regime. Customers in the network are indistinguishable. At node j there is an external Poisson–λ_j arrival stream, $\lambda_j \geq 0$. Customers arriving at node j from the outside or from inside of the network request for an amount of service time there which is exponentially distributed with mean 1. Service at node j is provided with intensity $\mu_j(n_j) > 0$, if there are $n_j > 0$ customers present at node j. (Note that the class of exponential FCFS single servers with state dependent service rates encompass exponential FCFS multiservers with state independent service rates.) All service and interarrival times are assumed to be independent.*

Movements of the customers in the network are governed by a Markovian routing mechanism: A customer on leaving node i selects with probability $r(i, j) \geq 0$ to visit node j next, and then enters node j immediately, commencing service if he finds the server idle, otherwise he joins the tail of the queue at node j; with probability $r(i, 0) \geq 0$ this customer decides to leave the network immediately ($\sum_{j=0}^{J} r(i, j) = 1$ for all $i \in \bar{J}$.) Given the departure node i the customer's routing decision is made independently of the network's history.

Let $\bar{J}_0 := \{0, 1, \ldots, J\}$. With $\lambda := \sum_{j=1}^{J} \lambda_j$, $r(0, j) := \frac{\lambda_j}{\lambda}$ and $r(0, 0) := 0$ we assume that the matrix $R := (r(i, j) : i, j \in \bar{J}_0)$ is irreducible.

Let $X_j(t)$ denote the number of customers present at node j at time $t \geq 0$, either waiting or in service (local queue length at node j), then $X(t) := (X_j(t) : j = 1, \ldots, J)$ is the joint queue length vector of the network at time t and $X = (X(t) : t \geq 0)$ is the joint queue length process of the Jackson network with state space $E := \mathbb{N}^J$.

Theorem 2.2 ([Jac57]) *The joint queue length process X of the Jackson network is a Markov process with transition rates $(q(x, y) : x, y \in E, x \neq y)$ given by:*
For $i, j \in \bar{J}$ and $x = (n_1, \ldots, n_J) \in E$

$$\begin{aligned} q(n_1, \ldots, n_i, \ldots, n_J;\, n_1, \ldots, n_i + 1, \ldots, n_J) &= \lambda_i \\ \text{and if} \quad n_i > 0 : & \\ q(n_1, \ldots, n_i, \ldots, n_J;\, n_1, \ldots, n_i - 1, \ldots, n_J) &= \mu_i(n_i)\, r(i, 0) \\ q(n_1, \ldots, n_i, \ldots, n_j, \ldots, n_J;\, n_1, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_J) &= \mu_i(n_i)\, r(i, j). \end{aligned}$$

Furthermore

$$q(x, x) = -\sum_{y \in E \setminus \{x\}} q(x, y) \text{ and } q(x, y) = 0 \text{ otherwise.}$$

The traffic equation of the network

$$\eta_j = \lambda_j + \sum_{i=1}^{J} \eta_i r(i,j), \quad j = 1, \ldots, J, \tag{1}$$

has a unique solution which we denote by $\eta = (\eta_1, \ldots, \eta_J)$.
We assume henceforth X *to be ergodic. The unique stationary and limiting distribution* π *of* X *is then*

$$\pi(n_1, \ldots, n_J) = K(J)^{-1} \prod_{j=1}^{J} \prod_{k=1}^{n_j} \left(\frac{\eta_j}{\mu_j(k)} \right), \quad (n_1, \ldots, n_J) \in \mathbb{N}^J$$

with normalization constant $K(J)$.

The steady state distribution of the closed analogue of the Jackson network was proved by Jackson some years later, see [Jac63], and rediscovered by Gordon and Newell [GN67]. We sketch the relevant results.

Definition 2.3 (Gordon-Newell network) *A Gordon-Newell network consists of a set of single server nodes* $\bar{J} := \{1, 2, \ldots, J\}$ *as described in the Definition 2.1 of the Jackson network without external arrivals and departures, i.e. the probabilities* $r(j,0)$ *for all* $j \in \bar{J}$ *and the total network arrival rate* λ *are set to zero. There are* $D > 0$ *indistinguishable customers cycling in the network according to an irreducible Markov matrix* $R = (r(i,j) : i, j \in \bar{J})$. *The service times are the same as in the open Jackson network case and the independence assumptions on service times and routing decisions are assumed to hold as well.*
Let $X_j(t)$ *denote the number of customers present at node* j *at time* $t \geq 0$, *either waiting or in service (local queue length at node* j*). Then* $X(t) := (X_j(t) : j \in \bar{J})$ *is the joint queue length vector of the network at time* t. *We denote by* $X = (X(t) : t \geq 0)$ *the joint queue length process of the Gordon-Newell network with state space* $S(J,D) := \{(n_1, n_2, \ldots, n_J) \in \mathbb{N}^J : n_1 + n_2 + \cdots + n_J = D\}$.

Theorem 2.4 ([Jac63], [GN67]) *The joint queue length process of the Gordon-Newell network* X *is a Markov process with transition rates* $(q(x,y) : x, y \in S(J,D), x \neq y)$ *given by:*
For $i, j \in \bar{J}$ *and* $x = (n_1, \ldots, n_J) \in S(J,D)$:

$$q(n_1 \ldots, n_i, \ldots, n_j, \ldots, n_J;\ n_1, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_J) = \mu_i(n_i) r(i,j) \ \textit{if}\ n_i > 0$$

and

$$q(x,x) = - \sum_{y \in S(D,J) \setminus \{x\}} q(x,y),$$
$$q(x,y) = 0, \ \textit{otherwise.}$$

The process X *is ergodic. Let* $\eta = (\eta_1, \ldots, \eta_J)$ *denote the unique probability solution of the traffic equation*

$$\eta_j = \sum_{i=1}^{J} \eta_i r(i,j), \ j \in \bar{J}. \tag{2}$$

The unique stationary and limiting distribution $\pi = \pi(J,D)$ *of* X *on* $S(J,D)$ *is*

$$\pi(n_1, \ldots, n_J) = G(D,J)^{-1} \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right) \ \textit{for}\ (n_1, \ldots, n_J) \in S(D,J), \tag{3}$$

with $G(D,J)$ *as normalization constant.*

BCMP networks are much more general than the so far presented exponential networks. In particular they allow different classes of customers and new service disciplines while the stationary solution is retained of product form. We consider networks with stochastic routing as in [BCMP75]. These can be transformed into networks with deterministic routing as used by Kelly [Kel79], see e.g. [Mel82]. Our description of BCMP networks follows [GP87], Chapter 4.

Definition 2.5 (BCMP network) *A BCMP network consists of J nodes with single or multi servers and $R \geq 1$ distinguishable classes of customers.*

Routing behaviour. *Customers belong to a customer class $r \in \bar{R} := \{1, 2, \ldots, R\}$ while awaiting or receiving service at a service center $i \in \bar{J} := \{1, 2, \ldots, J\}$ but may change their class when departing from a service center. Changing customer class and service center is done according to probabilities which are given by way of Markovian rules, fixed by transition probabilities*

$$\mathcal{P} = \big(p(ir, jr') : 0 \leq i, j \leq J,\ 1 \leq r, r' \leq R\big)\,.$$

$p(ir, jr')$ is the probability that a customer of class r at station i goes after completion of its service there to node j and turns into a member of the customer class r'.
The network may be closed or open. In the former case the indices i and j in the routing matrix $\mathcal{P}$ are restricted to $\bar{J}$. In the latter case the arrival process of new customers in the network is Poisson with state-independent rate λ. A new arrival is directed to node $i \in \bar{J}$ and turns to a member of customer class $r' \in \bar{R}$ with probability $\sum_{r \in \bar{R}} p(0r, ir')$ and a customer of class $r \in \bar{R}$ departing from node $i \in \bar{J}$ leaves the network with probability $\sum_{r' \in \bar{R}} p(ir, 0r')$. We set $p(0r, 0r') := 0$ for all $r, r' \in \bar{R}$, so that newly arriving customers do not immediately leave the network.

Service time distributions. *The service time distribution at node $i \in \bar{J}$ is a node and customer type dependent Cox distribution, i.e. the server consists of a set of exponential stages in line. This means: If at station i a customer of class $r \in \bar{R}$ enters the server from the waiting room, he immediately leaves it with probability $1 - b_0$ and with probability b_0 he receives a first service which is exponential distributed with mean μ_{ir1}. After its completion the customer's total service at station i expires with probability $(1 - b_1)$ or with probability b_1 he enters the second stage of the server where another service starts which is exponentially distributed with mean μ_{ir2}. After being served there he leaves the station with probability $(1 - b_2)$ or enters the third stage of the server with probability b_2. This procedure stops at the latest when the service at the last stage in line, l_{ir}, is completed. So μ_{irm} is the mean of the service time at the mth service stage of station i for a customer of class r and we denote by A_{irm} the probability that a customer of class r reaches the mth stage at station i, $A_{irm} = \prod_{h=0}^{m-1} b_h$ for $1 \leq m \leq l_{ir}$.*
Note that, e. g., exponential, hyperexponential and Erlang distributions are Cox distributions.

Service disciplines. *The service stations can obey any of the four following types:*

1. FIRST COME, FIRST SERVED (FCFS) *at a single server node. Then the service time needs to be (node dependent) exponentially distributed with the same mean for all classes of customers. The actual service rates at such station $i \in \bar{J}$ may depend on the local queue length and will be denoted by $\mu_i(n_i)$ if there are n_i customers in the station.*

2. PROCESSOR SHARING (PS), *the service time distributions can be type and node dependent Cox distributions. A customer at the station i receives $1/n_i$ seconds of service per second if there are n_i customers at the station $i \in \bar{J}$. So all customers receive in parallel an equal portion of their respective service time.*

3. INFINITE SERVER (IS), *i.e., the number of servers at the station is sufficiently large to house any number of customers. It follows that any new customer entering the station starts its service immediately. The service time distributions are type and node dependent Cox distributions.*

4. LAST COME, FIRST SERVED - PREEMPTIVE RESUME (LCFS - PR) *at a single server node. Newly arriving customers preempt ongoing service immediately. The preempted customers stay at the head of the queue and their service is resumed where it was interrupted, when the customer who caused the interruption has completed its service. Service time distributions are type and node dependent Cox distributions.*

Service times and routing decisions are independent of each other and of the past. We denote the solution of the traffic equation of this network by e_{ir} :

$$e_{ir} = \sum_{r'=1}^{R} \left(\sum_{j=1}^{J} e_{jr'} p(jr', ir) + p(0r', ir) \right) \text{ for } i \in \bar{J} \text{ and } r \in \bar{R}. \quad (4)$$

The state process of the network is $X = (X(t) : t \geq 0)$ *where* $X(t) = (x_1, x_2, \ldots, x_J)$ *is the state of the network at time* t. *The form of* x_i *depends on the type of service station at node* $i \in \bar{J}$: *If there are* n_i *customers at the station* i, *then for*

- *type 1:* $x_i = (c_{i1}, \ldots, c_{in_i})$, *where* c_{ij} *is the class of the* j*th customer waiting at station* i *in first come, first served order.*
- *type 2 or 3:* $x_i = ((c_{i1}, s_{i1}), \ldots, (c_{in_i}, s_{in_i}))$, *where* c_{ij} *is the class of the* j*th customer waiting at station* i *in the order of arrival and* s_{ij} *is its stage of the Cox model.*
- *type 4:* $x_i = ((c_{i1}, s_{i1}), \ldots, (c_{in_i}, s_{in_i}))$, *where* c_{ij} *is the class of the* j*th customer waiting at station* i *in last come, first served order and* s_{ij} *is its stage of the Cox model.*

F *denotes the state space of* X, k_{irm} *is the number of customers of class* r *at the* m*th stage at station* i, *and* $n_i := \sum_{r=1}^{R} \sum_{m=1}^{l_{ir}} k_{irm}$ *the total number of customers at station* i.

Theorem 2.6 ([BCMP75]) *The network process* $X = (X(t) : t \geq 0)$ *is a Markov process. We assume* X *to be ergodic on* F. *The unique stationary and limiting distribution of* X *is*

$$\pi(x_1, x_2, \ldots, x_J) = \Gamma^{-1} d(N) \prod_{i=1}^{J} f_i(x_i), \quad (x_1, x_2, \ldots, x_J) \in F \quad (5)$$

with the normalization constant $\Gamma = \sum_{(x_1, \ldots, x_J) \in F} d(N) \prod_{i=1}^{J} f_i(x_i)$,

$$N := \sum_{i=1}^{J} n_i, \qquad d(N) = \begin{cases} \lambda^N, & \text{if the network is open,} \\ 1, & \text{if the network is closed} \end{cases}$$

$$\text{and } f_i(x_i) = \begin{cases} \prod_{j=1}^{n_i} \frac{e_{ic_{ij}}}{\mu_i(j)}, & \text{if } i \text{ is of type 1,} \\ n_i! \prod_{r=1}^{R} \prod_{m=1}^{l_{ir}} \frac{(e_{ir} A_{irm} \mu_{irm}^{-1})^{k_{irm}}}{k_{irm}!}, & \text{if } i \text{ is of type 2,} \\ \prod_{r=1}^{R} \prod_{m=1}^{l_{ir}} \frac{(e_{ir} A_{irm} \mu_{irm}^{-1})^{k_{irm}}}{k_{irm}!}, & \text{if } i \text{ is of type 3,} \\ \prod_{j=1}^{n_i} e_{ic_{ij}} A_{ic_{ij} s_{ij}} \mu_{ic_{ij} s_{ij}}^{-1}, & \text{if } i \text{ is of type 4.} \end{cases}$$

3 Introducing breakdowns

For the networks described in Section 2 we consider the servers at the nodes to be unreliable, i.e., the nodes may break down. The breakdown event may occur in different ways: Nodes may break down as an isolated event or in groups simultaneously, and the repair of the nodes may end for each node individually or in groups as well. It is not required that those nodes which stopped service simultaneously return to service at the same time instant. To describe the system's evolution we have to enlarge the state space for each network process as will be described below.

Control of breakdowns and repairs is as follows:

Let $\bar{K} \subset \bar{J}$ be the set of nodes in down status and $\bar{I} \subset \bar{J} \backslash \bar{K}, \bar{I} \neq \emptyset$, be some subset of nodes in up status. Then the nodes of $\bar{I}$ break down concurrently with intensity $\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K} \cup \bar{I})$, if there are n_i customers at node i, $i \in \bar{K} \cup \bar{I}$. So breakdown of node set $\bar{I}$ depends on local loads of those nodes.
Nodes in down status neither accept new customers nor continue serving the old customers which will wait for the server's return. (At nodes i under repair the service intensities $\mu_i(n_i)$ are set to 0.) We will describe three regimes to handle routing connected with nodes in down status below.

Assume the nodes in $\bar{K}$ are under repair, $\bar{K} \subset \bar{J}, \bar{K} \neq \emptyset$. Then if $\bar{I} \subset \bar{K}, \bar{I} \neq \emptyset$, the nodes of $\bar{I}$ return from repair as a batch group with intensity $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K})$ and immediately resume their services. Note that at the breakdown instant there have been n_i customers at node i, $i \in \bar{K}$, which still are waiting there. Routing then has to be updated again as will be described below.
Intensities $\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K} \cup \bar{I})$ and $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{K})$ for occurrence of breakdowns and repairs have to be set under constraints. We found versatile rules for rather general classes of suitable intensities.

Definition 3.1 *The intensities for breakdowns, resp. repairs for $\bar{I} \neq \emptyset$ are defined by*

$$\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J}) := \frac{A(n_i : i \in \bar{K} \cup \bar{I})}{A(n_i : i \in \bar{K})}, \quad \textit{resp.} \quad \beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J}) := \frac{B(n_i : i \in \bar{K})}{B(n_i : i \in \bar{K} \backslash \bar{I})}, \tag{6}$$

where A and B are functions, $A, B : \cup_{l \in \bar{J}_0} \mathbb{N}^l \to \mathbb{R}_+ = [0, \infty)$.
Since $\alpha_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ and $\beta_{\bar{I}}^{\bar{K}}(n_i : i \in \bar{J})$ will serve as transition intensities for a Markovian system process they will be assumed henceforth to be finite.
This induces for example

$$A(n_i : i \in \bar{K}) = 0 \implies A(n_i : i \in \bar{K} \cup \bar{I}) = 0 \textit{ for all } \bar{K} \subset \bar{J}, \bar{I} \subset \bar{J} \backslash \bar{K}, \bar{I} \neq \emptyset,$$

which means that if at a fixed customer state $(n_1, \ldots, n_J) \in \mathbb{N}^J$ the common breakdown rate for a set of nodes, $\bar{K} \in \bar{J}$, is zero, than this set $\bar{K}$ cannot break down simultanously together with other nodes either.
For standardization we set $A(n_i : i \in \emptyset) := 1 =: B(n_i : i \in \emptyset)$.

It follows, e.g., that for a network in complete up status (all nodes are functioning) the breakdown intensity for node set $\bar{I} \subset \bar{J}, \bar{I} \neq \emptyset$ is $\alpha_{\bar{I}}^{\emptyset}(n_i : i \in \bar{I}) = A(n_i : i \in \bar{I})$ and the return intensity for the complete set of broken down nodes, say $\bar{K}$, is: $\beta_{\bar{K}}^{\bar{K}}(n_i : i \in \bar{K}) = B(n_i : i \in \bar{K})$.

Example 3.2 (Independent breakdown and repair) *An important case is when the rate for breakdown and the repair intensity of a single node, say i, depend only on the local load, n_i, at this node. The functions A and B are then assumed to be of product form*

$$A(n_i : i \in \bar{I}) = \prod_{i \in \bar{I}} a_i(n_i), \quad B(n_i : i \in \bar{I}) = \prod_{i \in \bar{I}} b_i(n_i),$$

which for all combinations of nodes $\bar K \subset \bar J$ leads to

$$\alpha_{\bar I}^{\bar K}(n_i : i \in \bar J) = \prod_{i \in \bar I} a_i(n_i) \quad \text{with } \bar I \cap \bar K = \emptyset \quad \text{and} \quad \beta_{\bar I}^{\bar K}(n_i : i \in \bar J) = \prod_{i \in \bar I} b_i(n_i) \quad \text{with } \bar I \subset \bar K.$$

Example 3.3 *The class of breakdown intensities in Definition 3.1 includes the following important special cases:*
(a) *Breakdown and repair intensities are independent of the customer load in the network. Then the functions A and B only depend on the node set concerned by the breakdown, resp. repair:*

$$A(n_i : i \in \bar I) = A(\bar I), \quad B(n_i : i \in \bar I) = B(\bar I) \ \text{ for all } \bar I \subset \bar J, \bar I \neq \emptyset.$$

With this assumptions the breakdown/repair process is Markovian on the state space $\mathcal{P}(\bar J)$ of all subsets of $\bar J$. The state description of this process can be equivalently chosen to be $\{0,1\}^J$. Then considering breakdowns as births and repairs as deaths we are in the range of multidimensional birth–death (b–d) processes on $\{0,1\}^J$ and the special intensities selected for this example coincide with those in [Ser93][(5), Table 1].
(b) *A more sophisticated regime is when nodes can only break down when they are idle and therefore the function A is defined as*

$$A(n_i : i \in \bar I) = A(\bar I) \prod_{i \in \bar I} \delta_{0\, n_i} \ \text{ for all } \bar I \subset \bar J, \bar I \neq \emptyset,$$

and it follows that the function B may take positive values only for (subsets of) nodes which are empty as well:

$$B(n_i : i \in \bar I) = B(\bar I) \prod_{i \in \bar I} \delta_{0\, n_i} \ \text{ for all } \bar I \subset \bar J, \bar I \neq \emptyset.$$

The behaviour of this system can be considered as a network of queues where empty servers take vacations. Such systems for the case of single nodes have found much interest in the literature during the last years, see [Dos90] for a review and [LAP01] for more recent developments.

Remark 3.4 *In the former case of Example 3.3 the distribution of the time until the first breakdown, when starting in a network state in which all nodes are working, is distributed according to the* MULTIVARIATE EXPONENTIAL MARSHALL-OLKIN DISTRIBUTION, *see [MO67].*

4 The steady states for the degradable networks

There are various strategies in the literature to handle breakdowns in case of unreliable machines or transmission channels. Either customers may be lost when the server breaks down, or they are stored at the waiting room, or they are rerouted to functioning nodes to obtain service there. Often these principles are translated into abstract stay–on and rerouting schemes. We adapt to our framework some abstract rules which are usually used to resolve blocking situations.
The subsequent theorems show that in case of unreliable servers we so obtain separable networks. We first present the probability measures of typical product form. These are the central features for our further development because they will serve as steady states for various network processes.

Definition 4.1 (Product form equilibrium states for degradable exponential networks) *For describing the breakdown of nodes in Jackson resp. Gordon-Newell networks we have to attach to the state spaces $E := \mathbb{N}^J$ resp. $E := S(J, D)$ of the corresponding network processes X an* *additional component which carries information of the reliability behaviour of the system described by a process Y. We introduce states of the form*

$$(\bar{I};\, n_1, n_2, \ldots, n_J) \in \mathcal{P}(\bar{J}) \times \mathbb{N}^J \text{ resp. } \mathcal{P}(\bar{J}) \times S(J, D) \text{ with } \bar{I} \neq \emptyset.$$

The meaning of such a prototype state is:
$\bar{I}$ is the set of nodes under repair. For $j \in \bar{J} \backslash \bar{I}$, the numbers $n_j \in \mathbb{N}$ indicate that at nodes j which work in a normal up status, there are n_j customers present; for $i \in \bar{I}$ the numbers $n_i \in \mathbb{N}$ indicate that at each node i which is in down status there are n_i customers that wait at node i for the return of the repaired server. Collecting these states we define for the networks new Markov processes $\widetilde{X} = (Y, X)$ on

$$\widetilde{E} = E \cup \big(\mathcal{P}(\bar{J}) \backslash \emptyset\big) \times \mathbb{N}^J \text{ resp. } \widetilde{E} = S(J, D) \cup \big(\mathcal{P}(\bar{J}) \backslash \emptyset\big) \times S(J, D).$$

On these state spaces $\widetilde{E}$ we define the probability measures of product form

$$\pi(n_1, \ldots, n_J) = C^{-1} \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right) \text{ with } (n_1, \ldots, n_J) \in E, \tag{7}$$

$$\pi(\bar{I};\, n_1, n_2, \ldots, n_J) = C^{-1} \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})} \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right), \tag{8}$$

with $(\bar{I};\, n_1, \ldots, n_J) \in \mathcal{P}(\bar{J}) \times \mathbb{N}^J$, $\bar{I} \neq \emptyset$.
Here η_j is the solution of the traffic equation (1) resp. (2) of the standard Jackson resp. Gordon-Newell network, C the normalization constant:

$$C = \sum_{(n_1, \ldots, n_J) \in E} \left[\prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)} \right) \left(1 + \sum_{\substack{\bar{I} \subset \bar{J} \\ \bar{I} \neq \emptyset}} \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})} \right) \right].$$

We assume C to be finite.

For a Jackson resp. Gordon-Newell network according to Definition 2.1 resp. Definition 2.3, with control of breakdowns according to Definition 3.1 and with suitably defined routing processes (see the descriptions in Chapters 5 and 6 below) this probability measures will serve as uniquely defined stationary and limiting distributions.

Definition 4.2 (Product form equilibrium states for degradable BCMP networks)
For introducing breakdowns into BCMP networks we enlarge the state space F of the describing network process X by defining new states of the form

$$(\bar{I};\, x_1, \ldots, x_J) \in \mathcal{P}(\bar{J}) \times F \text{ with } \bar{I} \neq \emptyset$$

with the meaning that the nodes of set $\bar{I}$ are under repair and x_i for $i \in \bar{J} \backslash \bar{I}$ describe the customer situation at node $i, i \in \bar{J} \backslash \bar{I}$, as defined in Definition 2.5, while the x_i for $i \in \bar{I}$ indicate the customer situation at node $i, i \in \bar{I}$, which will not change until node i is repaired and resumes its service. Collecting these states we define for the degradable BCMP network a new Markov process $\widetilde{X}$ on the state space $\widetilde{F} := F \cup \big(\mathcal{P}(\bar{J}) \backslash \emptyset\big) \times F$.

We define the probability measure of product form for the network process $\widetilde{X} = (Y, X)$ *on* $\widetilde{F}$ *by*

$$\pi\,(x_1, x_2, \ldots, x_J) = \Lambda^{-1}\, d(N) \prod_{i=1}^{J} f_i(x_i) \ \textit{with}\ (x_1, \ldots, x_J) \in F \tag{9}$$

$$\pi\left(\bar{I};\, x_1, x_2, \ldots, x_J\right) = \Lambda^{-1}\, \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})}\, d(N) \prod_{i=1}^{J} f_i(x_i) \tag{10}$$

with $(\bar{I};\, x_1, \ldots, x_J) \in \widetilde{F}$, $\bar{I} \neq \emptyset$ *and with* n_i *denoting the number of customers at node* i, $i \in \bar{J}$. *(Recall that* $d(N) = d(\sum_{i=1}^{J} n_i)$.*) Here*

$$\Lambda = \sum_{(x_1, \ldots, x_J) \in F} d(N) \prod_{i=1}^{J} f_i(x_i) \left(1 + \sum_{\substack{\bar{I} \subset \bar{J} \\ \bar{I} \neq \emptyset}} \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})}\right)$$

serves as normalization constant, which we assume to be finite, and the functions d *and* $f_i, i \in \bar{I}$, *are defined in Theorem 2.6.*
In $\widetilde{X} = (Y, X)$ *the component* Y *carries information about the reliability status of the network, while* X *is the usual performance evaluation vector with information about the load of the system.*

For a BCMP network according to Definition 2.5 with the control of breakdowns according to Definition 3.1 and with suitably defined routing processes (see the descriptions in Chapters 5 and 6 below) this probability measure will serve as uniquely defined stationary and limiting distribution.

5 Routing in case of breakdowns: Blocking

We shall now discuss several routing protocols in case of breakdowns of servers. The first one is derived from a well known principle used to resolve blocking situations in networks with resource constraints.
Blocking mechanisms have been developed for stochastic networks, in which a customer cannot leave a node after having obtained his service there because the population in his next destination node has reached the capacity of waiting plus service places there. Therefore the customer cannot join this node and the flow through the network according to the routing matrix is delayed or totally stopped. We give a short survey about the most prominent methods to model this blocking problem. For more detailed reviews of blocking techniques see [Onv90], [Per90] or the recently appeared book [BNPO01].

Definition 5.1 (Blocking Mechanisms)
• Blocking after service (bas)
After completing his service at node i *a customer choses his next destination node* j *according to his routing instruction. If at that moment node* j *is not able to accept further customers, the blocked customer is forced to occupy server* i *until the situation at node* j *has changed and he can enter there. Node* i *is blocked during this waiting period, i.e. it cannot begin the service of other customers, who might be waiting in its waiting room. If several nodes are simultaneously blocked by the same node* j*, then the "first blocked - first unblocked- rule" usually determines the order in which the nodes will be unblocked when a departure occurs from the destination node* j.
• Blocking before service (bbs)
A customer at node i *selects his next destination node* j *according to his routing rule before he starts receiving service at node* i. *If node* j *is full, node* i *immediately becomes blocked. When a departure occurs from node* j*, node* i *becomes unblocked and service begins. Every time when the*

destination node j becomes full again during that service at i, the service there has to be stopped and node i becomes blocked again. Depending upon whether the customer at the blocked node i is allowed to occupy the position in front of the server or not, one can distinguish the two cases BBS-SERVER OCCUPIED *and* BBS-SERVER NOT OCCUPIED. *Of course this is only meaningful when node i has finite waiting capacity.*

• REPETITIVE SERVICE (RS) *A customer after being served at node i choses his next destination node j according to his routing instruction. If he finds this node j full, he stays at node i to obtain another service. When this additional service expires he either selects his destination node anew according to his routing rule* (RS - RANDOM DESTINATION) *or the previously chosen node j remains his fixed destination and the customer's service at node i has to be repeated until at the end of a service node j is able to receive another customer* (RS - FIXED DESTINATION).

5.1 The Blocking principle RS–RD

We have selected the blocking principle REPETITIVE SERVICE – RANDOM DESTINATION (RS–RD) to obtain a new routing regime for our networks in cases that there are nodes under repair which cannot continue their services. For that we forbid customers to enter these nodes and redirect them to other nodes in the network so that we have to deal with a situation like blocking. The principle RS–RD has the advantage that the service intensity at a blocked node formally must not be changed and is independent of the situation at other nodes. We derive a new routing matrix according to RS–RD for our networks which regulates the customers' moves between the nodes that are not under repair.

We adapt the REPETITIVE SERVICE – RANDOM DESTINATION (RS–RD) concept to our setting of Jackson resp. Gordon-Newell networks as follows.

Definition 5.2 (RS–RD with reversible routing in exponential networks) *Assume the nodes in $\bar{I}$ are the nodes of the Jackson resp. Gordon-Newell network presently under repair. Then the routing probabilities are redefined on $\bar{J}_0\backslash\bar{I}$ according to*

$$\tilde{r}^{\bar{I}}(i,j) = \begin{cases} r(i,j), & i,j \in \bar{J}_0\backslash\bar{I},\ i \neq j, \\ r(i,i) + \sum_{k\in\bar{I}} r(i,k), & i \in \bar{J}_0\backslash\bar{I},\ i=j, \\ 0, & \textit{otherwise}. \end{cases} \tag{11}$$

In the case of a Jackson network it follows for the external arrival rates

$$\tilde{\lambda}_j^{\bar{I}} = \lambda\tilde{r}^{\bar{I}}(0,j) = \lambda r(0,j) = \lambda_j \textit{ for nodes } j \in \bar{J}\backslash\bar{I} \textit{ functioning in up status,}$$
$$\tilde{\lambda}_j^{\bar{I}} = \lambda\tilde{r}^{\bar{I}}(0,j) = 0 \textit{ for nodes under repair } j \in \bar{I}\ (j \neq 0).$$

Note that external arrivals may be rejected with positive probability to an immediate departure, because arrivals to nodes under repair are rerouted:

$$\tilde{r}^{\bar{I}}(0,0) = r(0,0) + \sum_{k\in\bar{I}} r(0,k) = \sum_{k\in\bar{I}} \frac{\lambda_k}{\lambda} \geq 0.$$

We assume further that the routing matrix of the original process, see Definition 2.1 (resp. Definition 2.3), is reversible:

$$\eta_j \cdot r(j,i) = \eta_i \cdot r(i,j) \textit{ for } i,j \in \bar{J}_0 \textit{ (resp. } i,j \in \bar{J}) \quad \textit{with } \eta_j \textit{ from (1) (resp. (2))} \tag{12}$$

Applying this routing regime results in product form theorems:

Theorem 5.3 *A Jackson network according to Definition 2.1 with reversible routing and with breakdown and repair intensities from Definition 3.1 and rerouting according to* RS–RD *(Definition 5.2) has a stationary distribution of product form given by (7) and (8).*

Proof: The described network possesses the following stationary equations:

$$
\begin{aligned}
&\pi(n_1,\ldots,n_J)\left(\sum_{j=1}^{J}(1-\delta_{0n_j})\mu_j(n_j)(1-r(j,j))+\sum_{j=1}^{J}\lambda_j+\sum_{\substack{\bar I\subset\bar J\\ \bar I\neq\emptyset}}A(n_i:i\in\bar I)\right)\\
&=\sum_{j=1}^{J}(1-\delta_{0n_j})\pi(n_1,\ldots,n_j-1,\ldots,n_J)\lambda_j\\
&+\sum_{j=1}^{J}(1-\delta_{0n_j})\sum_{\substack{i=1\\ i\neq j}}^{J}\pi(n_1,\ldots,n_i+1,\ldots,n_j-1,\ldots,n_J)\mu_i(n_i+1)r(i,j)\\
&+\sum_{j=1}^{J}\pi(n_1,\ldots,n_j+1,\ldots,n_J)\mu_j(n_j+1)r(j,0)\\
&+\sum_{\substack{\bar I\subset\bar J\\ \bar I\neq\emptyset}}\pi(\bar I;\, n_1,n_2,\ldots,n_J)B(n_i:i\in\bar I)\quad\text{ for all }(n_1\ldots,n_J)\in\mathbb{N}^J,
\end{aligned}\tag{13}
$$

$$
\begin{aligned}
&\pi(\bar I;\, n_1,n_2,\ldots,n_J)\Bigg(\sum_{j\in\bar J\setminus\bar I}(1-\delta_{0n_j})\mu_j(n_j)(1-\tilde r^{\bar I}(j,j))+\sum_{j\in\bar J\setminus\bar I}\tilde\lambda_j^{\bar I}\\
&\qquad\qquad+\sum_{\substack{\bar K\subset\bar I\\ \bar K\neq\emptyset}}\frac{B(n_i:i\in\bar I)}{B(n_i:i\in\bar I\setminus\bar K)}+\sum_{\substack{\bar K\subset\bar J\setminus\bar I\\ \bar K\neq\emptyset}}\frac{A(n_i:i\in\bar I\cup\bar K)}{A(n_i:i\in\bar I)}\Bigg)\\
&=\sum_{j\in\bar J\setminus\bar I}(1-\delta_{0n_j})\pi(\bar I;\, n_1,n_2,\ldots,n_j-1,\ldots,n_J)\tilde\lambda_j^{\bar I}\\
&+\sum_{j\in\bar J\setminus\bar I}(1-\delta_{0n_j})\sum_{\substack{k\in\bar J\setminus\bar I\\ k\neq j}}\pi(\bar I;\, n_1,\ldots,n_k+1,\ldots,n_j-1,\ldots,n_J)\mu_k(n_k+1)\tilde r^{\bar I}(k,j)\\
&+\sum_{j\in\bar J\setminus\bar I}\pi(\bar I;\, n_1,n_2,\ldots,n_j+1,\ldots,n_J)\mu_j(n_j+1)\tilde r^{\bar I}(j,0)\\
&+\sum_{\substack{\bar K\subset\bar I\\ \bar K\neq\emptyset}}\pi(\bar I\setminus\bar K;\, n_1,n_2,\ldots,n_J)\frac{A(n_i:i\in\bar I)}{A(n_i:i\in\bar I\setminus\bar K)}\\
&+\sum_{\substack{\bar K\subset\bar J\setminus\bar I\\ \bar K\neq\emptyset}}\pi(\bar I\cup\bar K;\, n_1,n_2,\ldots,n_J)\frac{B(n_i:i\in\bar I\cup\bar K)}{B(n_i:i\in\bar I)}
\end{aligned}\tag{14}
$$

for all $(\bar I;\, n_1,n_2,\ldots,n_J)\in\mathcal{P}(\bar J)\times\mathbb{N}^J$ with $\bar I\neq\emptyset$.

We have to show that the distributions (7) and (8) solve these equations (13) and (14). In the first equality (13) the term $\pi(n_1,\ldots,n_J)A(n_i:i\in\bar I)$ at the left corresponds to the term $\pi(\bar I;\, n_1,\ldots,n_J)B(n_i:i\in\bar I)$ at the right side for each $\bar I\subset\bar J,\bar I\neq\emptyset$. Then exactly the global

stationary equation of the standard Jackson network with J nodes is left over. It is solved by the stationary distribution (7).
In the second equation (14) the term at the left side

$$\pi(\bar{I};\, n_1, n_2, \ldots, n_J)\frac{B(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I}\backslash\bar{K})} = \pi(n_1, \ldots, n_J)\frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I}\backslash\bar{K})}$$

corresponds to the term at the right side

$$\pi(\bar{I}\backslash\bar{K};\, n_1, \ldots, n_J)\frac{A(n_i : i \in \bar{I})}{A(n_i : i \in \bar{I}\backslash\bar{K})} = \pi(n_1 \ldots, n_J)\frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I}\backslash\bar{K})}$$

for each $\bar{K} \subset \bar{I}, \bar{K} \neq \emptyset$, just as the term

$$\pi(\bar{I};\, n_1, \ldots, n_J)\frac{A(n_i : i \in \bar{I} \cup \bar{K})}{A(n_i : i \in \bar{I})} = \pi(n_1 \ldots, n_J)\frac{A(n_i : i \in \bar{I} \cup \bar{K})}{B(n_i : i \in \bar{I})}$$

at the left side corresponds for each $\bar{K} \subset \bar{J}\backslash\bar{I}, \bar{K} \neq \emptyset$ to the term

$$\pi(\bar{I} \cup \bar{K};\, n_1, \ldots, n_J)\frac{B(n_i : i \in \bar{I} \cup \bar{K})}{B(n_i : i \in \bar{I})} = \pi(n_1 \ldots, n_J)\frac{A(n_i : i \in \bar{I} \cup \bar{K})}{B(n_i : i \in \bar{I})}$$

at the right equation side. The remainder here is the global stationary equation of the standard Jackson network with node set $\bar{J}\backslash\bar{I}$, service rates μ_j, external arrival rates $\tilde{\lambda}_j^{\bar{I}}$ and routing probabilities $\tilde{r}^{\bar{I}}(i,j)$ for $i,j \in \bar{J}\backslash\bar{I}$. It is solved by

$$\pi(\bar{I};\, n_1, \ldots, n_J) = \prod_{j \in \bar{J}\backslash\bar{I}} \prod_{i=1}^{n_j} \left(\frac{\tilde{\eta}_j^{\bar{I}}}{\mu_j(i)}\right) \tilde{C}^{-1}(n_i : i \in \bar{I}).$$

In this expression in connection with the remainder equation the parameters n_i for $i \in \bar{I}$ are constants and $\tilde{\eta}^{\bar{I}}$ is the solution of the traffic equation

$$\tilde{\eta}_j^{\bar{I}} = \tilde{\lambda}_j^{\bar{I}} + \sum_{i \in \bar{J}\backslash\bar{I}} \tilde{\eta}_i^{\bar{I}} \tilde{r}^{\bar{I}}(i,j), \quad j \in \bar{J}\backslash\bar{I}.$$

Therefore it only remains to show that

$$\tilde{\eta}_j^{\bar{I}} = \eta_j \text{ for all } j \in \bar{J}\backslash\bar{I}. \tag{15}$$

By using the definition of the new routing probabilities $\tilde{r}^{\bar{I}}(i,j)$ from (11) this identity can be verified:

$$\begin{aligned}
&\tilde{\lambda}_j^{\bar{I}} + \sum_{i \in \bar{J}\backslash\bar{I}} \eta_i \tilde{r}^{\bar{I}}(i,j) \\
&= \lambda_j + \sum_{\substack{i \in \bar{J}\backslash\bar{I} \\ i \neq j}} \eta_i r(i,j) + \eta_j \Big(r(j,j) + \sum_{i \in \bar{I}} r(j,i)\Big) \\
&= \lambda_j + \sum_{i \in \bar{J}\backslash\bar{I}} \eta_i r(i,j) + \sum_{i \in \bar{I}} \eta_j r(j,i) \\
&= \eta_j - \sum_{i \in \bar{I}} \eta_i r(i,j) + \sum_{i \in \bar{I}} \eta_i r(i,j) = \eta_j,\ j \in \bar{J}\backslash\bar{I},
\end{aligned}$$

where the last but one equality follows from the premise of reversibility. Finally we define

$$\tilde{C}^{-1}(n_i : i \in \bar{I}) = \prod_{j\in\bar{I}} \prod_{i=1}^{n_j} \left(\frac{\eta_j}{\mu_j(i)}\right) C^{-1} \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})},$$

to obtain the proposed form for $\pi(\bar{I};\, n_1, \ldots, n_J)$. ■

Theorem 5.4 *A Gordon-Newell network according to Definition 2.3 with reversible routing and with breakdown and repair intensities from Definition 3.1 and rerouting according to* RS–RD *(Definition 5.2) has a stationary distribution of product form given by (7) and (8).*

Proof:
The proof of Theorem 5.3 is modified by setting $\lambda = 0$ and $r(j, 0) = 0$ for all $j \in \bar{J}$. ■

We now adapt the Repetitive Service – Random Destination (RS–RD) concept to the setting of BCMP networks:

Definition 5.5 (RS–RD with reversible routing in BCMP networks) *Assume the nodes in $\bar{I}$ are the nodes of the BCMP network presently under repair. Then the routing probabilities are restricted to nodes with indices from $\bar{J}_0\backslash\bar{I}$ according to*

$$\tilde{p}^{\bar{I}}(ir, jr') = \begin{cases} p(ir, jr'), & i, j \in \bar{J}_0\backslash\bar{I},\ i \neq j,\ r, r' \in \bar{R}, \\ p(ir, ir') + \sum_{k\in\bar{I}} p(ir, kr'), & i \in \bar{J}_0\backslash\bar{I},\ i = j,\ r, r' \in \bar{R}, \\ 0, & \textit{otherwise}. \end{cases} \tag{16}$$

In the case of an open network it follows for the external arrival probabilities

$$\tilde{p}^{\bar{I}}(0r, jr') = p(0r, jr') \textit{ for nodes } j \in \bar{J}\backslash\bar{I} \textit{ functioning in up status, } r, r' \in \bar{R},$$
$$\tilde{p}^{\bar{I}}(0r, jr') = 0 \textit{ for nodes under repair } j \in \bar{I}\ (j \neq 0),\ r, r' \in \bar{R}.$$

Note that external arrivals may be rerouted with positive probability to an immediate departure

$$\tilde{p}^{\bar{I}}(0r, 0r') = p(0r, 0r') + \sum_{k\in\bar{I}} p(0r, kr')\ \geq 0.$$

We assume further that the routing matrix of the original process, see Definition 2.5, is reversible in the node component when the change in customer class remains the same, that means:

$$e_{ir'} \cdot p(ir', jr) = e_{jr'} \cdot p(jr', ir) \textit{ for } i, j \in \bar{J}_0 \textit{ and } r, r' \in \bar{R} \quad \textit{with } e_{ir} \textit{ from } (4). \tag{17}$$

The main result in this section follows.

Theorem 5.6 *A BCMP network from Definition 2.5 with reversible routing according to* (17) *and with breakdown and repair intensities from Definition 3.1 and rerouting according to* RS–RD *(Definition 5.5) has a stationary distribution of product form given by* (9) *and* (10).

Proof.
The proof uses properties of the time-reversed Markov process $\bar{X}$ of $\tilde{X}$, given by $\bar{X}(t) := \tilde{X}(-t), t \in \mathbb{R}$, which can be found in the Appendix. We apply Lemma A.3 to our framework in order to prove that the stationary distribution π of the enlarged network process $\tilde{X}$ on the state space

$\widetilde{F} = F \cup \left(\mathcal{P}(\bar{J})\backslash\emptyset\right) \times F$ is given by the probabilities (9) and (10) of Definition 4.2 . Denote the Q-matrix of the process $\widetilde{X}$ by $\widetilde{Q} = \left(\widetilde{q}(z,z') : z, z' \in \tilde{F}\right)$.
First notice that the numbers (9) and (10) sum to one by definition of Λ.

For each $\bar{H} \subset \bar{J}$ consider a (standard) BCMP network with customer classes $\{1, 2, \ldots, R\} =: \bar{R}$ and node set $\bar{J}$ with the restriction that at nodes in $\bar{H} \subset \bar{J}$ some state $x|_{\bar{H}} = (x_i : i \in \bar{H})$ is prescribed and frozen, and that customers only move between nodes in the set $\bar{J}\backslash\bar{H}$ according to the routing matrix $\tilde{\mathcal{P}}^{\bar{H}} = (\tilde{p}^{\bar{H}}(ir, jr') : i, j \in \bar{J}\backslash\bar{H}, r, r' \in \bar{R})$ from Definition 5.5. The Q-matrix $Q^{\bar{J}\backslash\bar{H}} = \left(q^{\bar{J}\backslash\bar{H}}(z,z') : z, z' \in F\right)$ of the corresponding Markov process $X^{\bar{J}\backslash\bar{H}}$, which is in general not irreducible on the state space F, has only entries $q^{\bar{J}\backslash\bar{H}}(z,z')$ greater than zero for $z, z' \in F, z \neq z'$, if $z|_{\bar{H}} = z'|_{\bar{H}} = x|_{\bar{H}}$. In the latter case we have $q^{\bar{J}\backslash\bar{H}}(z,z') = \widetilde{q}((\bar{H}, z), (\bar{H}, z'))$.

The stationary distribution of the process $X^{\bar{J}\backslash\bar{H}}$ is according to Theorem 2.6 for $(x_1, x_2, \ldots, x_J) \in F$ given by

$$\pi^{\bar{H}}(x_i : i \in \bar{J}) = \Upsilon_{\bar{H}}^{-1}(x_i : i \in \bar{H})\, d\left(\sum_{i \in \bar{J}\backslash\bar{H}} n_i\right) \prod_{i \in \bar{J}\backslash\bar{H}} f_i(x_i), \quad \bar{H} \subset \bar{J} \tag{18}$$

with $\Upsilon_{\bar{H}}(x_i : i \in \bar{H})$ as normalization constant, which may depend on the frozen state of subnetwork $\bar{H}$, and with x_i, functions d and f_i defined in Theorem 2.6, but the latter in a first step with $e_{ir}^{\bar{H}}$, $i \in \bar{J}\backslash\bar{H}, r \in \bar{R}$, as the solutions of the corresponding traffic equations,

$$e_{ir}^{\bar{H}} = \sum_{r'=1}^{R}\left(\tilde{p}^{\bar{H}}(0r', ir) + \sum_{j \in \bar{J}\backslash\bar{H}} e_{jr'}^{\bar{H}}\, \tilde{p}^{\bar{H}}(jr', ir)\right), i \in \bar{J}\backslash\bar{H}, r \in \bar{R}, \tag{19}$$

instead of the numbers e_{ir}, $i \in \bar{J}, r \in \bar{R}$ from (4).
These solutions $e_{ir}^{\bar{H}}$, $i \in \bar{J}\backslash\bar{H}, r \in \bar{R}$, can be shown to equal those $e_{ir}, i \in \bar{J}\backslash\bar{H}, r \in \bar{R}$, as follows:

$$\begin{aligned}
&\sum_{r'=1}^{R}\left(\tilde{p}^{\bar{H}}(0r', ir) + \sum_{j=1}^{J} e_{jr'}\tilde{p}^{\bar{H}}(jr', ir)\right) \\
&= \sum_{r'=1}^{R}\left(p(0r', ir) + \sum_{\substack{j \in \bar{J}\backslash\bar{H} \\ j \neq i}} e_{jr'}p(jr', ir) + e_{ir'}\left(p(ir', ir) + \sum_{k \in \bar{H}} p(ir', kr)\right)\right) \\
&= \sum_{r'=1}^{R}\left(p(0r', ir) + \sum_{j \in \bar{J}\backslash\bar{H}} e_{jr'}p(jr', ir) + \sum_{k \in \bar{H}} e_{ir'}p(ir', kr)\right) \\
&= \sum_{r'=1}^{R}\left(p(0r', ir) + \sum_{j \in \bar{J}\backslash\bar{H}} e_{jr'}p(jr', ir) + \sum_{k \in \bar{H}} e_{kr'}p(kr', ir)\right) \\
&= \sum_{r'=1}^{R}\left(p(0r', ir) + \sum_{j \in \bar{J}} e_{jr'}p(jr', ir)\right) = e_{ir},
\end{aligned}$$

where the last but two equality follows from reversibility (17).
So the stationary distribution of the process $X^{\bar{J}\backslash\bar{H}}$ with essential node set $\bar{J}\backslash\bar{H}$ and routing matrix $\tilde{\mathcal{P}}^{\bar{H}} = (\tilde{p}^{\bar{H}}(ir, jr') : i, j \in \bar{J}\backslash\bar{H}, r, r' \in \bar{R})$ equals up to a constant the stationary distribution of the

network process X of a (standard) BCMP network. The constant will be choosen depending on the frozen local states $x_i, i \in \bar{H}$.
We define a matrix $\bar{Q} = (\bar{q}(z,z') : z, z' \in \widetilde{F})$ by $\bar{q}(z) := q(z)$ for all $z \in \tilde{F}$ and for $z \neq z'$ by

$$\bar{q}(z,z') = \begin{cases} \frac{\pi^{\emptyset}(z')\, q^{\bar{J}\setminus\emptyset}(z',z)}{\pi^{\emptyset}(z)}, & \text{if } z, z' \in F \text{ with } \pi^{\emptyset} \text{ according to (18)}, \\[2ex] \frac{\pi^{\bar{H}}(z'|_F)\, q^{\bar{J}\setminus\bar{H}}(z'|_F, z|_F)}{\pi^{\bar{H}}(z|_F)}, & \text{if } z, z' \in \{\bar{H}\} \times F \text{ with } z|_{\bar{H}} = z'|_{\bar{H}} \text{ for } \bar{H} \in \mathcal{P}(\bar{J}), \bar{H} \neq \emptyset \\ & \text{and with } \pi^{\bar{H}} \text{ according to (18)}, \\[2ex] \frac{B(n_i : i \in \bar{H})}{B(n_i : i \in \bar{I})}, & \text{if } z \in \{\bar{H}\} \times F \text{ and } z' \in \{\bar{I}\} \times F \text{ with } z|_F = z'|_F \\ & \text{for } \bar{H}, \bar{I} \in \mathcal{P}(\bar{J}), \bar{I} \subset \bar{H},\ \bar{I} \neq \bar{H}, \\[2ex] \frac{A(n_i : i \in \bar{I})}{A(n_i : i \in \bar{H})}, & \text{if } z \in \{\bar{H}\} \times F \text{ and } z' \in \{\bar{I}\} \times F \text{ with } z|_F = z'|_F \\ & \text{for } \bar{H}, \bar{I} \in \mathcal{P}(\bar{J}), \bar{H} \subset \bar{I},\ \bar{I} \neq \bar{H}, \\[2ex] 0, & \text{otherwise}. \end{cases} \tag{20}$$

It remains to show according to (37) that the numbers $\bar{q}(z,z')$ for $z, z' \in \widetilde{F}$, $z \neq z'$ fulfill the equation

$$\pi(z')\widetilde{q}(z',z) = \pi(z)\bar{q}(z,z') \text{ for } z, z' \in \widetilde{F}, \tag{21}$$

with π from Definition 4.2. In case $z \in \{\bar{H}\} \times F$ and $z' \in \{\bar{I}\} \times F$ with $z|_F = z'|_F$ for $\bar{H}, \bar{I} \in \mathcal{P}(\bar{J})$, $\bar{I} \neq \bar{H}$, this follows at once from the Definition (20), because of

$$\widetilde{q}(z',z) = \frac{A(n_i : i \in \bar{H})}{A(n_i : i \in \bar{I})}, \text{ if } \bar{I} \subset \bar{H} \text{ and } \widetilde{q}(z',z) = \frac{B(n_i : i \in \bar{I})}{B(n_i : i \in \bar{H})}, \text{ if } \bar{H} \subset \bar{I}.$$

We give explicitly only the equations in case $\bar{I} \subset \bar{H}$:

$$\begin{aligned} \pi(z')\widetilde{q}(z',z) &= \Lambda^{-1} \frac{A(n_i : i \in \bar{I})}{B(n_i : i \in \bar{I})}\, d(N) \prod_{i=1}^{J} f_i(x_i)\, \frac{A(n_i : i \in \bar{H})}{A(n_i : i \in \bar{I})} \\ &= \Lambda^{-1} \frac{A(n_i : i \in \bar{H})}{B(n_i : i \in \bar{I})}\, d(N) \prod_{i=1}^{J} f_i(x_i) \\ &= \Lambda^{-1} \frac{A(n_i : i \in \bar{H})}{B(n_i : i \in \bar{H})}\, d(N) \prod_{i=1}^{J} f_i(x_i)\, \frac{B(n_i : i \in \bar{H})}{B(n_i : i \in \bar{I})} = \pi(z)\bar{q}(z,z'). \end{aligned}$$

In cases $z, z' \in \{\bar{H}\} \times F$ with $z|_{\bar{H}} = z'|_{\bar{H}}$ for $\bar{H} \in \mathcal{P}(J), \bar{H} \neq \emptyset$ (as well as analogous in case $z, z' \in F$) the equation (21) follows immediately with the help of Lemma A.2: We have defined $\bar{q}(z,z')$ as entries of the Q-matrix of the time-reversed process assigned to the describing process $X^{\bar{J}\setminus\bar{H}}$ of a BCMP network with node set $\bar{J}\setminus\bar{H}$, routing according to Definition 5.5 and Q-matrix $Q^{\bar{J}\setminus\bar{H}}$. As the stationary distribution of this process equals up to the constant $\Upsilon_{\bar{H}}^{-1}(x_i : i \in \bar{H})$ the stationary distribution of the process $X^{\emptyset} = X$, we define

$$\Upsilon_{\bar{H}}^{-1}(x_i : i \in \bar{H}) := \Lambda^{-1} \frac{A(n_i :\ i \in \bar{H})}{B(n_i :\ i \in \bar{H})}\, d(\sum_{i \in \bar{H}} n_i) \prod_{i \in \bar{H}} f_i(x_i)$$

with Λ^{-1} from Definition 4.2. Then (21) immediately follows from

$$\widetilde{q}(z,z') = q^{\bar{J}\setminus\bar{H}}(z|_F,\, z'|_F) \text{ for all } z,z' \in \{\bar{H}\}\times F,\ z\neq z' \text{ with } z|_{\bar{H}} = z'|_{\bar{H}}. \qquad \blacksquare$$

The requirement of having reversible routing for the original network without unreliability phenomenon cannot be expected to be removed in the setting of this subsection. Even in the simpler case of introducing control of blocking this is a standard assumption necessary to obtain explicit expressions for the steady state distribution. It is an advantage of the routing control mechanisms of the next sections that such an assumption is not needed to obtain simple to use equilibria.

5.2 The Stalling principle

The next principle to handle breakdowns is borrowed again from how to deal with blocking, see e.g. [Mik88], [Mit74], [Dij93](Subsection 3.6.2). It can be considered to be an extension of the BLOCKING BEFORE SERVICE (BBS)- principle. When some node j breaks down and enters its repair status the service is not only stopped at those nodes where a customer with next destination node j is in service but at all the other nodes as well. Therefore the total movement of customers in the network is brought to a stop. In addition customers who want to enter the network from the outside are rerouted to an immediate departure. The service at all nodes is not resumed until all broken down nodes are repaired again. Following [Sch84] we call this principle STALLING.
We assume that servers in up status at nodes where the service is interrupted due to a stalling situation behave as being in a *hot stand–by* mode. This means that these nodes can break down although working capacity is only hold ready and not effectively used. Further the breakdown rates of these nodes or collections of them are the same as if they are working on the frozen population size there.

Definition 5.7 (STALLING) *If there is any breakdown of either a singleton node or a group of nodes, then all arrival streams to the network and all service processes at the nodes in up status are completely interrupted and resumed only when all nodes are repaired again.*

With this rerouting principle we obtain the same stationary behaviour as in the case of the RS-RD mechanism.

Theorem 5.8 *Consider a Jackson network according to Definition 2.1 with breakdown and repair intensities from Definition 3.1 and rerouting according to* STALLING *(Definition 5.7). Nodes being in up status but not serving because of lying dormant under the stalling regime are in hot stand–by mode, i.e. the breakdown rates are not changed by the stalling regime. Then this Jackson network has a stationary distribution of product form given by (7) and (8).*

Proof:
With arguments analogous to those in the proof of Theorem 5.3 it can be shown that the stationary distribution from (7) and (8) fulfills the global balance equations for the network process:

$$\pi(n_1,\ldots,n_J)\left(\sum_{j=1}^{J}(1-\delta_{0n_j})\mu_j(n_j)(1-r(j,j))+\sum_{j=1}^{J}\lambda_j+\sum_{\substack{\bar I\subset\bar J\\ \bar I\neq\emptyset}}A(n_i:i\in\bar I)\right)$$
$$=\sum_{j=1}^{J}(1-\delta_{0n_j})\pi(n_1,\ldots,n_j-1,\ldots,n_J)\lambda_j$$
$$+\sum_{j=1}^{J}(1-\delta_{0n_j})\sum_{\substack{i=1\\ i\neq j}}^{J}\pi(n_1,\ldots,n_i+1,\ldots,n_j-1,\ldots,n_J)\mu_i(n_i+1)r(i,j)$$
$$+\sum_{j=1}^{J}\pi(n_1,\ldots,n_j+1,\ldots,n_J)\mu_j(n_j+1)r(j,0)$$
$$+\sum_{\substack{\bar I\subset\bar J\\ \bar I\neq\emptyset}}\pi(\bar I;\, n_1,n_2,\ldots,n_J)B(n_i:i\in\bar I)\quad\text{for all } (n_1\ldots,n_J)\in\mathbb{N}^J,$$

$$\pi(\bar I;\, n_1,n_2,\ldots,n_J)\left(\sum_{\substack{\bar K\subset\bar I\\ \bar K\neq\emptyset}}\frac{B(n_i:i\in\bar I)}{B(n_i:i\in\bar I\backslash\bar K)}+\sum_{\substack{\bar K\subset\bar J\backslash\bar I\\ \bar K\neq\emptyset}}\frac{A(n_i:i\in\bar I\cup\bar K)}{A(n_i:i\in\bar I)}\right)$$
$$=\sum_{\substack{\bar K\subset\bar I\\ \bar K\neq\emptyset}}\pi(\bar I\backslash\bar K;\, n_1,n_2,\ldots,n_J)\frac{A(n_i:i\in\bar I)}{A(n_i:i\in\bar I\backslash\bar K)}\tag{22}$$
$$+\sum_{\substack{\bar K\subset\bar J\backslash\bar I\\ \bar K\neq\emptyset}}\pi(\bar I\cup\bar K;\, n_1,n_2,\ldots,n_J)\frac{B(n_i:i\in\bar I\cup\bar K)}{B(n_i:i\in\bar I)}$$
$$\text{for all } (\bar I;\, n_1,n_2,\ldots,n_J)\in\mathcal{P}(\bar J)\times\mathbb{N}^J \text{ with } \bar I\neq\emptyset.$$

■

Note that the occurrence of (22) indicates formally the consequences of the hot stand–by regime for nodes in up status which are interrupted during their services due to breakdowns of other servers: No server active and no arrivals but servers may break down nevertheless.

Theorem 5.9 *Consider a Gordon-Newell network according to Definition 2.3 with breakdown and repair intensities from Definition 3.1 and rerouting according to* STALLING *(Definition 5.7). Nodes being in up status but not serving because of lying dormant under the stalling regime are in hot stand–by mode, i.e. the breakdown rates are not changed by the stalling regime. Then this Gordon-Newell network has a stationary distribution of product form given by (7) and (8).*

Proof:
The proof of Theorem 5.8 is modified by setting $\lambda = 0$ and $r(j,0) = 0$ for all $j \in \bar J$. ■

Theorem 5.10 *Consider a BCMP network according to Definition 2.5 with breakdown and repair intensities from Definition 3.1 and rerouting according to* STALLING *(Definition 5.7). Nodes being in up status but not serving because of lying dormant under the stalling regime are in hot stand–by mode, i.e. the breakdown rates are not changed by the stalling regime. Then this BCMP network has a stationary distribution of product form given by* (9) *and* (10).

Proof:
With arguments analogous to those in the proof of Theorem 5.6 the theorem can be proved again with the help of Lemma A.3. ■

6 Routing in case of breakdowns: Skipping

The third principle to handle breakdowns relies on methods from the theory of taboo probabilities for Markov chains. These principles seem to be introduced into queueing network theory independently several times and were used to resolve blocking, see e.g.[EF98] and [Ser99][Chapter 3.6], (where it is called blocking and rerouting) and the references therein. As a general methodology this was already introduced by Schassberger [Sch84] and used to generally modify queueing network processes without destroying product form of the stationary distribution. Later on, based on Schassberger's result, it was used in [DS96] to construct general abstract network processes.

For a first understanding of the skipping principle consider the routing process of the Jackson network from Definition 2.1 as an irreducible, homogenous Markov chain $V = (V_n : n \in \mathbb{N})$ on state space $\bar{J}_0 := \{0, 1, \ldots, J\}$ with transition matrix $\mathcal{R} = (r(i,j) : i, j \in \bar{J}_0)$. Assume that travelling customers in the network are no longer allowed to enter node set $\bar{I} \subset \bar{J}$ ($\bar{I} \neq \emptyset, \bar{I} \neq \bar{J}$), i.e., $\bar{I}$ is a so-called taboo set for the process V that has to be skipped. Graphically spoken, the skipping rule for the customers' jumps under this restriction is:
If the customer present at node $i \in \bar{J}_0 \backslash \bar{I}$ selects (with probability $r(i,j)$) for his jump's destination node $j \in \bar{J}_0 \backslash \bar{I}$, this jump is allowed and immediately performed and the customer joins node j for service. If (with probability $r(i,k)$) he selects for his jump's destination node $k \in \bar{I}$, he only performs an imaginary jump to that node, spending no time there, but jumping on immediately according to the routing matrix $\mathcal{R}$, i.e. with probability $r(k,l)$ he selects another possible successor node l; if $l \in \bar{J}_0 \backslash \bar{I}$ then the jump is performed and the customer joins node l for service, but if $l \in \bar{I}$ the customer has to perform another random choice as if he would depart from l; and so on. It is easily seen that such behaviour would also be obtained by setting the service intensities at the nodes under repair to ∞.

To formalize this, for an irreducible, homogenous Markov chain V on a countable state space Z with transition probability matrix $S = (s(z,z') : z, z' \in Z)$ we define the restriction of V to the state space $Z_0 \subseteq Z$, such that the taboo set $Z \backslash Z_0$ is finite, as a homogenous Markov chain $W = (W_n : n \in \mathbb{N})$ as follows:

$$W_0 \equiv V_0 \quad \text{and with } h(0,\omega) \equiv 0$$

$$W_n(\omega) := V_{h(n,\omega)}(\omega) \text{ with } h(n,\omega) := \inf \{m \in \mathbb{N} : m > h(n-1,\omega), V_m(\omega) \in Z_0\} \text{ for } n \geq 1.$$

Note that the process W may start in the taboo set $Z \backslash Z_0$.
Let $s_0(z,z') := \mathcal{P}(W_1 = z' | W_0 = z)$ be the probability that the process V enters the allowed state space Z_0 for the first time after $n = 0$ at state $z' \in Z_0$, when starting in $z \in Z$. The matrix $\mathcal{S}_0 = (s_0(z,z') : z \in Z, z' \in Z_0)$ is stochastic because of

$$\sum_{z' \in Z_0} s_0(z,z') = \sum_{z' \in Z_0} \mathcal{P}(W_1 = z' | W_0 = z) = 1.$$

With the aid of the First Entrance Method and the strong Markov property we derive from the probability $s_0(z,z')$ for $z \in Z$ and $z' \in Z_0$ the *skipping principle* (23):

$$s_0(z,z') = s(z,z') + \sum_{z'' \in Z \backslash Z_0} s(z,z'') s_0(z'',z'). \tag{23}$$

This follows from

$$
\begin{aligned}
s_0(z,z') &= \mathcal{P}(W_1 = z' | W_0 = z) \\
&= \mathcal{P}(W_1 = z', V_1 \in Z_0 | W_0 = z) + \mathcal{P}(W_1 = z', V_1 \in Z\backslash Z_0 | W_0 = z) \\
&= \mathcal{P}(V_1 = z' | V_0 = z) + \sum_{z'' \in Z\backslash Z_0} \mathcal{P}(V_1 = z'' | W_0 = z) \\
&\qquad\qquad\qquad\qquad\qquad \cdot \mathcal{P}(W_1 = z' | W_0 = z, V_1 = z'') \\
&= s(z,z') + \sum_{z'' \in Z\backslash Z_0} \mathcal{P}(V_1 = z'' | V_0 = z) \cdot \underbrace{\mathcal{P}(W_1 = z' | V_1 = z'')}_{=\mathcal{P}(W_1 = z' | W_0 = z'')} \\
&= s(z,z') + \sum_{z'' \in Z\backslash Z_0} s(z,z'') s_0(z'',z') \text{ for } z \in Z \text{ and } z' \in Z_0.
\end{aligned}
$$

It is essential that invariant measures of the process V are invariant measures for $S_0|_{(Z_0 \times Z_0)}$, i.e. of $W|_{Z_0}$ as well:

Lemma 6.1 ([Sch84]) *A vector $u = (u(z) : z \in Z)$ with $u\mathcal{S} = u$ also fulfills $u|_{Z_0}\mathcal{S}_0|_{(Z_0 \times Z_0)} = u|_{Z_0}$ with $u|_{Z_0} = (u(z) : z \in Z_0)$.*

Proof: Applying the skipping principle (23) yields for all $z \in Z$ and $z' \in Z_0$:

$$u(z)s_0(z,z') = u(z)s(z,z') + u(z) \sum_{z'' \in Z\backslash Z_0} s(z,z'')s_0(z'',z').$$

Summation over all $z \in Z$ produces for all $z' \in Z_0$:

$$
\begin{aligned}
&\Rightarrow \sum_{z \in Z} u(z)s_0(z,z') = \underbrace{\sum_{z \in Z} u(z)s(z,z')}_{=u(z')} + \sum_{z'' \in Z\backslash Z_0} s_0(z'',z') \underbrace{\sum_{z \in Z} u(z)s(z,z'')}_{=u(z'')} \\
&\Leftrightarrow \sum_{z \in Z} u(z)s_0(z,z') = u(z') + \sum_{z \in Z\backslash Z_0} u(z)s_0(z,z') \\
&\Leftrightarrow u(z') = \sum_{z \in Z_0} u(z)s_0(z,z').
\end{aligned}
$$

■

By expressing the SKIPPING PRINCIPLE (23) in terms of matrices, we obtain an explicit description of the full skipping matrix $S_0 = (s_0(z,z') : z \in Z, z' \in Z_0)$, as it will now be developed. We rewrite (23) and obtain similarly

$$S_0 = S|_{(Z \times Z_0)} + S|_{(Z \times Z\backslash Z_0)} \cdot S_0|_{(Z\backslash Z_0 \times Z_0)}, \tag{24}$$

Restricting this on $(Z\backslash Z_0 \times Z_0)$ we obtain for the last factor an equation which can be iterated.

$$S_0|_{(Z\backslash Z_0 \times Z_0)} = S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot S_0|_{(Z\backslash Z_0 \times Z_0)}. \tag{25}$$

Iterating (25) we obtain

$$
\begin{aligned}
&S_0|_{(Z\backslash Z_0 \times Z_0)} = S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot S_0|_{(Z\backslash Z_0 \times Z_0)} \\
&= S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot \left(S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot S_0|_{(Z\backslash Z_0 \times Z_0)}\right)
\end{aligned}
$$

This suggests formally:

Lemma 6.2 *The* SKIPPING PRINCIPLE (23) *is explicitly solved by the following transition probability matrices:*

$$S_0|_{(Z\backslash Z_0 \times Z_0)} = \sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n \cdot S|_{(Z\backslash Z_0 \times Z_0)} \tag{26}$$

$$S_0 = S|_{(Z \times Z_0)} + S|_{(Z \times Z\backslash Z_0)} \cdot \left(\sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n \right) \cdot S|_{(Z\backslash Z_0 \times Z_0)}. \tag{27}$$

Proof: With (26) equation (24) yields the expression (27) for the full skipping matrix S_0. Inserting (26) into (25) proves formally formula (26):

$$\begin{aligned} S_0|_{(Z\backslash Z_o \times Z_0)} &= S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot S_0|_{(Z\backslash Z_0 \times Z_0)} \\ &= S|_{(Z\backslash Z_0 \times Z_0)} + S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} \cdot \sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n \cdot S|_{(Z\backslash Z_0 \times Z_0)} \\ &= S|_{(Z\backslash Z_0 \times Z_0)} + \sum_{n=1}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n \cdot S|_{(Z\backslash Z_0 \times Z_0)} \\ &= \sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n \cdot S|_{(Z\backslash Z_0 \times Z_0)}. \end{aligned}$$

It remains to show that the series $\sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n$ converges:
The matrix $S|_{(Z\backslash Z_0 \times Z\backslash Z_0)} =: T = (t(z,z') : z, z' \in Z\backslash Z_0)$ is substochastic and in generally not irreducible on $Z\backslash Z_0 \times Z\backslash Z_0$. But because the jump matrix $S = (s(z,z') : z, z' \in Z)$ is irreducible and stochastic, it follows, that for each state $z_0 \in Z\backslash Z_0$ there exists a state $w = w(z_0) \in Z\backslash Z_0$ which is attainable from z_0 in a finite number of steps and which has a direct connection to the state space Z_0 :
Therefore for all $z_0 \in Z\backslash Z_0$ there exist $w(z_0) \in Z\backslash Z_0$ and $h(z_0) \in \mathbb{N}$ so that $t^{h(z_0)}(z_0, w(z_0)) > 0$ and $\sum_{x \in Z\backslash Z_0} t(w(z_0), x) < 1$. It follows for all $z_0 \in Z\backslash Z_0$:

$$\begin{aligned} \sum_{x \in Z\backslash Z_0} t^{h(z_0)+1}(z_0, x) &= \sum_{x \in Z\backslash Z_0} \sum_{y \in Z\backslash Z_0} t^{h(z_0)}(z_0, y)\, t(y, x) \\ &= \sum_{y \in Z\backslash Z_0} t^{h(z_0)}(z_0, y) \underbrace{\sum_{x \in Z\backslash Z_0} t(y, x)}_{<1 \text{ for } y = w(z_0)} \quad < 1, \end{aligned}$$

because of $t^{h(z_0)}(z_0, w(z_0)) > 0$. With $H := \sup_{z_0 \in Z\backslash Z_0} (h(z_0) + 1)$ we obtain for the matrix norm

$$||T^H|| := \sup_{z_0 \in Z\backslash Z_0} \sum_{x \in Z\backslash Z_0} t^H(z_0, x) < 1$$

It follows:

$$\lim_{N \to \infty} || \sum_{n=0}^{N} T^n || \leq \lim_{N \to \infty} \sum_{n=0}^{N} ||T^H||^{\frac{n}{H}} < \infty,$$

so that the series $\sum_{n=0}^{\infty} (S|_{(Z\backslash Z_0 \times Z\backslash Z_0)})^n$ converges geometrically. ■

Definition 6.3 (SKIPPING) *The skipping principle* (23) *applied to the routing process with taboo set $\bar{I}$ yields in Jackson (resp. Gordon-Newell) networks new Markov routing matrices*

$$\hat{\mathcal{R}}^{\bar{I}} = \left(\hat{r}^{\bar{I}}(i,j) : i, j \in \bar{J}_0 \backslash \bar{I} \text{ (resp. } \bar{J} \backslash \bar{I}) \right)$$

which are determined by the following set of equations:

$$\hat{r}^{\bar{I}}(j,k) = r(j,k) + \sum_{i\in\bar{I}} r(j,i)\hat{r}^{\bar{I}}(i,k), \quad k,j \in \bar{J}_0\backslash\bar{I} \ (resp.\ \bar{J}\backslash\bar{I}), \tag{28}$$

$$\hat{r}^{\bar{I}}(i,k) = r(i,k) + \sum_{l\in\bar{I}} r(i,l)\hat{r}^{\bar{I}}(l,k), \quad i \in \bar{I}, k \in \bar{J}_0\backslash\bar{I} \ (resp.\ \bar{J}\backslash\bar{I},) \tag{29}$$

and $\hat{r}^{\bar{I}}(i,k) = 0$, *otherwise.*

It follows for the new arrival rates $\hat{\lambda}_j^{\bar{I}}$ *for* $j \in \bar{J}$ *in Jackson networks:*

$$\hat{\lambda}_j^{\bar{I}} = \lambda\hat{r}^{\bar{I}}(0,j) = \underbrace{\lambda r(0,j)}_{=\lambda_j} + \sum_{i\in\bar{I}} \underbrace{\lambda r(0,i)}_{=\lambda_i} \hat{r}^{\bar{I}}(i,j) \ for\ j \in \bar{J}\backslash\bar{I} \quad and \quad \hat{\lambda}_k^{\bar{I}} = 0 \ for\ k \in \bar{I}. \tag{30}$$

The main results in the framework of exponential networks elaborate on the generalized product form probabilities from Definition 4.1 again.

Theorem 6.4 *For a Jackson network according to Definition 2.1 with breakdown and repair intensities from Definition 3.1 and rerouting according to the* SKIPPING RULE *(Definition 6.3) the state process on* $\widetilde{E}$ *has a stationary distribution of product form given in (7) and (8).*

Proof: With arguments analogous to those in the proof of Theorem 5.3 it suffices to show that the solutions of the traffic equations of a standard Jackson network and of a Jackson network with node set $\bar{J}\backslash\bar{I}$, service rates $\mu_j(\cdot)$, external arrival rates $\hat{\lambda}_j^{\bar{I}}$ from (30) and routing probabilities $\hat{r}^{\bar{I}}(i,j)$ from (28) for $i,j \in \bar{J}\backslash\bar{I}$ are equal, i.e that $\hat{\eta}_j^{\bar{I}} = \eta_j$ for all $j \in \bar{J}\backslash\bar{I}$ with $\hat{\eta}_j^{\bar{I}}$ as the solution of

$$\hat{\eta}_j^{\bar{I}} = \hat{\lambda}_j^{\bar{I}} + \sum_{i\in\bar{J}\backslash\bar{I}} \hat{\eta}_i^{\bar{I}}\hat{r}^{\bar{I}}(i,j) \text{ for } j \in \bar{J}\backslash\bar{I}.$$

Recall the traffic equation (1) of the Jackson network

$$\eta_j = \lambda_j + \sum_{i\in\bar{J}} \eta_i r(i,j) \text{ for } j \in \bar{J}. \tag{31}$$

Summing up this equation (31) over all $j \in \bar{J}$ yields with the definition $r(0,0) = 0$:

$$\sum_{j\in\bar{J}} \eta_j = \sum_{j\in\bar{J}} \lambda_j + \sum_{j\in\bar{J}}\sum_{i\in\bar{J}} \eta_i r(i,j) = \lambda \underbrace{\sum_{j\in\bar{J}} r(0,j)}_{=1} + \sum_{i\in\bar{J}} \eta_i \underbrace{\sum_{j\in\bar{J}} r(i,j)}_{=1-r(i,0)}$$

$$\iff \lambda = \sum_{i\in\bar{J}} \eta_i r(i,0). \tag{32}$$

Summarizing equations (31) and (32) yields with $\lambda = \eta_0$:

$$\eta_j = \sum_{i\in\bar{J}_0} \eta_i r(i,j) \text{ for } j \in \bar{J}_0.$$

I.e. $\eta = (\eta_0, \eta_1, \ldots, \eta_J)$ is an invariant measure for the routing process according to the routing matrix $\mathcal{R}$. It follows from Lemma 6.1 that for all potential groups of taboo nodes, $\bar{I}$, the accompanying new routing matrix $\hat{\mathcal{R}}^{\bar{I}} = (\hat{r}^{\bar{I}}(i,j) : i,j \in \bar{J}_0\backslash\bar{I})$ fulfills: $\eta\hat{\mathcal{R}}^{\bar{I}} = \eta$, i.e., we have shown with $\hat{\lambda}_j^{\bar{I}}$ from (30):

$$\eta_j = \hat{\lambda}_j^{\bar{I}} + \sum_{i\in\bar{J}\backslash\bar{I}} \eta_i\hat{r}^{\bar{I}}(i,j) \text{ for } j \in \bar{J}\backslash\bar{I}$$

and this means $\hat{\eta}_j^{\bar{I}} = \eta_j$ for all $j \in \bar{J}\backslash\bar{I}$. ∎

Theorem 6.5 *For a Gordon-Newell network according to Definition 2.3 with breakdown and repair intensities from Definition 3.1 and rerouting according to the* SKIPPING RULE *(Definition 6.3) the state process on $\widetilde{E}$ has a stationary distribution of product form given in (7) and (8).*

Proof: The proof of Theorem 6.4 is modified by setting $\lambda = 0$ and $r(j,0) = 0$ for all $j \in \bar{J}$. ■

Application of the SKIPPING RULE (23) to the general BCMP networks has to take into consideration carefully the customer type changes.

Definition 6.6 (SKIPPING) *Applying the* SKIPPING PRINCIPLE (23) *in BCMP networks to the two-dimensional routing process governed by the routing matrix*

$$P = \big(p(ir, jr') : i, j \in \bar{J}_0,\ r, r' \in \bar{R}\big)$$

with taboo set $\bar{I} \times \bar{R}$ yields new Markov routing matrices

$$\hat{\mathcal{P}}^{\bar{I}} = \left(\hat{p}^{\bar{I}}(ir, jr') : i, j \in \bar{J}_0 \backslash \bar{I}, r, r' \in \bar{R}\right)$$

which are determined by the following set of equations:

$$\hat{p}^{\bar{I}}(ir, jr') = p(ir, jr') + \sum_{k \in \bar{I}} \sum_{r'' \in \bar{R}} p(ir, kr'')\, \hat{p}^{\bar{I}}(kr'', jr'), \quad i \in \bar{J}_0, j \in \bar{J}_0 \backslash \bar{I},\ r, r' \in \bar{R}, \tag{33}$$

and $\hat{p}^{\bar{I}}(ir, jr') = 0$, otherwise.

It follows for the new arrival probabilities $\hat{\lambda}_j^{\bar{I}}$ in case of an open network:

$$\hat{\lambda}_j^{\bar{I}} = \lambda \hat{r}^{\bar{I}}(0, j) = \underbrace{\lambda r(0,j)}_{=\lambda_j} + \sum_{i \in \bar{I}} \underbrace{\lambda r(0,i)}_{=\lambda_i} \hat{r}^{\bar{I}}(i,j) \text{ for } j \in \bar{J} \backslash \bar{I} \quad \text{and} \quad \hat{\lambda}_k^{\bar{I}} = 0 \text{ for } k \in \bar{I}. \tag{34}$$

Note that the graphical interpretation of the skipping scheme presented at the beginning of this section and according to (23) reveals that a customer while travelling (in zero time) through the taboo set $\bar{I} \times \bar{R}$ may experience type changes. The main result within this setting is now:

Theorem 6.7 *For a BCMP network according to Definition 2.5 with breakdown and repair intensities from Definition 3.1 and rerouting governed by the* SKIPPING REGIME *(Definition 6.6) the state process on $\widetilde{F}$ has a stationary distribution of product form given by* (9) *and* (10).

Proof: The proof proceeds analogously to the proof of Theorem 5.6. We skip details here and show only that the solutions of the traffic equations of a standard BCMP network with node set $\bar{J}$ and of a BCMP network with node set $\bar{J} \backslash \bar{I}$ and routing probabilities $\hat{p}^{\bar{I}}(ir, jr')$ for $i, j \in \bar{J} \backslash \bar{I},\ r, r' \in \bar{R}$ according to Definition 6.6 are equal, i.e $\hat{e}_{ir}^{\bar{I}} = e_{ir}$ for all $i \in \bar{J} \backslash \bar{I}$ with $\hat{e}_{ir}^{\bar{I}}$ as the solution to

$$\hat{e}_{ir}^{\bar{I}} = \sum_{r' \in \bar{R}} \left(\hat{p}^{\bar{I}}(0r', ir) + \sum_{j \in \bar{J} \backslash \bar{I}} \hat{e}_{jr'}^{\bar{I}} \hat{p}^{\bar{I}}(jr', ir) \right) \text{ for } i \in \bar{J} \backslash \bar{I},\ r \in \bar{R}$$

and with e_{ir}, $i \in \bar{J}, r \in \bar{R}$ from the traffic equation of the BCMP network (4). Multiplying the SKIPPING RULE for BCMP networks (33) with e_{ir} yields for $i \in \bar{J}, j \in \bar{J} \backslash \bar{I}, r, r' \in \bar{R}$:

$$e_{ir} \hat{p}^{\bar{I}}(ir, jr') = e_{ir} p(ir, jr') + \sum_{r'' \in \bar{R}} \sum_{k \in \bar{I}} e_{ir} p(ir, kr'') \hat{p}^{\bar{I}}(kr'', jr'). \tag{35}$$

Summing equation (35) over all $r \in \bar{R}$ and $i \in \bar{J}$ yields for all $j \in \bar{J}\backslash\bar{I}$ and $r' \in \bar{R}$:

$$\sum_{r\in\bar{R}}\sum_{i\in\bar{J}} e_{ir}\hat{p}^{\bar{I}}(ir,jr') = \underbrace{\sum_{r\in\bar{R}}\sum_{i\in\bar{J}} e_{ir}p(ir,jr')}_{e_{jr'}-\sum_{r\in\bar{R}}p(0r,jr')} + \sum_{r''\in\bar{R}}\sum_{k\in\bar{I}}\hat{p}^{\bar{I}}(kr'',jr')\underbrace{\sum_{r\in\bar{R}}\sum_{i\in\bar{J}} e_{ir}p(ir,kr'')}_{e_{kr''}-\sum_{r\in\bar{R}}p(0r,kr'')}$$

$$\Leftrightarrow \sum_{r\in\bar{R}}\sum_{i\in\bar{J}} e_{ir}\hat{p}^{\bar{I}}(ir,jr') = e_{jr'} - \sum_{r\in\bar{R}}p(0r,jr') + \sum_{r''\in\bar{R}}\sum_{k\in\bar{I}}\hat{p}^{\bar{I}}(kr'',jr')\left(e_{kr''} - \sum_{r\in\bar{R}}p(0r,kr'')\right)$$

$$\Leftrightarrow \sum_{r\in\bar{R}}\sum_{i\in\bar{J}\backslash\bar{I}} e_{ir}\hat{p}^{\bar{I}}(ir,jr') = e_{jr'} - \sum_{r\in\bar{R}}p(0r,jr') - \sum_{r''\in\bar{R}}\sum_{k\in\bar{I}}\sum_{r\in\bar{R}}p(0r,kr'')\hat{p}^{\bar{I}}(kr'',jr')$$

$$\Leftrightarrow e_{jr'} = \sum_{r\in\bar{R}}\sum_{i\in\bar{J}\backslash\bar{I}} e_{ir}\hat{p}^{\bar{I}}(ir,jr') + \sum_{r\in\bar{R}}\underbrace{\left(p(0r,jr') + \sum_{r''\in\bar{R}}\sum_{k\in\bar{I}}p(0r,kr'')\hat{p}^{\bar{I}}(kr'',jr')\right)}_{\hat{p}^{\bar{I}}(0r,jr')}$$

$$\Leftrightarrow e_{jr'} = \sum_{r\in\bar{R}}\sum_{i\in\bar{J}\backslash\bar{I}} e_{ir}\hat{p}^{\bar{I}}(ir,jr') + \sum_{r\in\bar{R}}\hat{p}^{\bar{I}}(0r,jr').$$

So the equality $e^{\bar{I}}_{jr'} = e_{jr'}$ is proved for all $j \in \bar{J}\backslash\bar{I}$ and $r' \in \bar{R}$. ■

7 Computation of availability and performance measures

The network processes $\widetilde{X} = (Y, X)$ with their resp. steady states given in Definitions 4.1 and 4.2 include as observables the information on availability (reliability properties) and queue lengths (performance properties). To be more precise: For each model $\widetilde{X}(t) = (Y(t), X(t))$ we see that from the very definition of the Markovian state we can determine reliability aspects and performance aspects simply by computing marginal distributions and their characteristics. This will be explained by BCMP network examples.

We consider BCMP networks with breakdown and repair intensities which are independent from the customer load in the network (see case **(a)** of Example 3.3). In this case we obtain simple and explicit expressions for the availability and performance measures. We denote by $|x_j|$ the number of customers at node j in the network if x_j is the local customer state at node j according to Definition 2.5.

Lemma 7.1 *In stationary degradable BCMP networks according to Definition 4.2 with breakdown and repair intensities, which only depend on the set of nodes concerned (see Example 3.3*(**a**)*), the probability $\pi(\bar{I})$ for simultaneous repair of node set $\bar{I} \subset \bar{J}, \bar{I} \neq \emptyset$, is given by*

$$\pi(\bar{I}) = \frac{1}{1 + \sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}} \cdot \frac{A(\bar{I})}{B(\bar{I})}$$

Proof: With the normalization constants Γ (from the standard BCMP network according to

Theorem 2.6) and Λ (from the degradable BCMP network according to Definition 4.2) we obtain:

$$\begin{aligned}
\pi(\bar{I}) &= \sum_{(x_1,\ldots,x_J)\in F} \pi(\bar{I};\, x_1,\ldots,x_J) = \Lambda^{-1}\,\frac{A(\bar{I})}{B(\bar{I})} \sum_{(x_1,\ldots,x_J)\in F} d(N) \prod_{i=1}^{J} f_i(x_i) \\
&= \frac{\Gamma}{\Lambda}\cdot\frac{A(\bar{I})}{B(\bar{I})} = \frac{1}{1+\sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}}\cdot\frac{A(\bar{I})}{B(\bar{I})}
\end{aligned}$$

since

$$\frac{\Gamma}{\Lambda} = \frac{\sum_{(x_1,\ldots,x_J)\in F} d(N)\prod_{i=1}^{J} f_i(x_i)}{\left(1+\sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}\right)\sum_{(x_1,\ldots,x_J)\in F} d(N)\prod_{i=1}^{J} f_i(x_i)} = \frac{1}{1+\sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}} \qquad \blacksquare \quad (36)$$

Lemma 7.2 *In stationary degradable BCMP networks according to Definition 4.2 with breakdown and repair intensities, which only depend on the set of nodes concerned (see Example 3.3*(**a**)*), the average queue length at node $j\in\bar{J}$, denoted by $\tilde{E}(X_j)$, is the same as the corresponding mean in the standard BCMP network, $E(X_j)$.*

Proof:

$$\begin{aligned}
\tilde{E}(X_j) &= \sum_{n_j=0}^{\infty} n_j \cdot \sum_{\substack{(x_1,\ldots,x_J)\in F:\\ |x_j|=n_j}} \Lambda^{-1}\left(1+\sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}\right) d(N)\prod_{i=1}^{J} f_i(x_i) \\
&= \left(1+\sum_{\substack{\bar{I}\subset\bar{J}\\ \bar{I}\neq\emptyset}} \frac{A(\bar{I})}{B(\bar{I})}\right)\Lambda^{-1} \underbrace{\sum_{n_j=0}^{\infty} n_j \cdot \sum_{\substack{(x_1,\ldots,x_J)\in F:\\ |x_j|=n_j}} d(N)\prod_{i=1}^{J} f_i(x_i)}_{=\Gamma\, E(X_j)} = E(X_j),
\end{aligned}$$

where the last equality follows from equation (36). $\blacksquare$

Summarizing the observations in Lemmas 7.1 and 7.2 we have shown that in case of breakdown and repair rates from Example 3.3 (**a**) the mean value analysis for availability and performance measures can be decoupled.

As remarked in Example 3.3 (**a**) the marginal process $Y=(Y(t):t\geq 0)$ which describes the breakdown/repair status of the nodes is Markov by itself. Therefore the result of Lemma 7.1 is not surprising because $\pi(\bar{I})$ is the steady state of the multidimensional birth–death process on $\{0,1\}^J$ with transition rates given there.

On the other hand Y and the joint queue length network process X are definitly dependent, because the transition mechanism of X depends on the actual state of Y. Therefore the result of Lemma 7.2 is less intuitive. In fact this result does not hold with more general system dynamics as can be seen from the next example, which shows, that in case of general breakdown and repair intensities the average queue lengths at the nodes in the degradable networks can differ from the corresponding means in the standard networks:

Example 7.3 *In a stationary degradable BCMP network according to Definition 4.2 with only one node which can only break down when the server is idle (see Example 3.3* (**b**)*), the average queue* *length $\tilde{E}(X_1)$ differs from the corresponding mean $E(X_1)$ in the network without breakdown, as can easily be computed:*

$$\tilde{E}(X_1) = \sum_{n_1=0}^{\infty} n_1 \cdot \Lambda^{-1} \left(1 + \frac{A(n_1)}{B(n_1)}\right) d(n_1) \sum_{\substack{x_1 \in F \\ |x_1|=n_1}} f_1(x_1) = \Lambda^{-1} \sum_{n_1=1}^{\infty} n_1 \cdot d(n_1) \sum_{\substack{x_1 \in F \\ |x_1|=n_1}} f_1(x_1)$$

$$= \frac{\sum_{n_1=1}^{\infty} n_1\, d(n_1) \sum_{\substack{x_1 \in F \\ |x_1|=n_1}} f_1(x_1)}{\sum_{x_1 \in F} \left(1 + \frac{A(|x_1|)}{B(|x_1|)}\right) d(|x_1|) f_1(x_1)} = \frac{\sum_{n_1=1}^{\infty} n_1\, d(n_1) \sum_{\substack{x_1 \in F \\ |x_1|=n_1}} f_1(x_1)}{\sum_{x_1 \in F} d(|x_1|) f_1(x_1) + \sum_{\substack{x_1 \in F \\ |x_1|=0}} \underbrace{d(0)}_{=1} \cdot \underbrace{f_1(x_1)}_{=1} \cdot \frac{A(0)}{B(0)}}$$

$$= E(X_1) \cdot \frac{\Gamma}{\Gamma + \frac{A(0)}{B(0)}} < E(X_1).$$

The queueing model of Example 7.3 can be termed a vacation model (see [Dos90] for a review and [LAP01] for more recent developments): Whenever the server becomes idle he waits at most an exponentially distributed time (with parameter $A(0)$) for an arrival of a new customer. If no customer arrives the server takes a vacation which is exponentially distributed with parameter $B(0)$. When a vacation is ongoing new customers are rejected. When the vacation expires the returned server again waits an exponential $A(0)$ time for new customers. If some new customer arrives service immediately commences. Otherwise a new vacation is taken by the server, and so on.

8 Conclusions

The survey in the paper [CM96] which we sketched in the introduction elucidates two principles for dealing with the hard problem of to incorporate reliability and performance aspects in a unified Markovian system process with the aim to obtain explicit and simply to use formulas in both directions.
First: Define a detailed Markov model and then use approximative solution procedures, usually decomposition methods.
Second: Find a simple Markov model with explicitly computable steady state distribution.
In the realm of the first principle are the described Poisson approximation, the MMPP approximation, and the joint–state approximation, while the reduced work–rate approximation belongs to the second principle.

The advantage of the class of models constructed according to the second principle is that after initially performing the modeling procedure during the evaluation process no additional adhoc assumptions for further approximations are needed.
But in case of the reduced workload approximation there is a severe drawback: In the modeling process we loose the breakdown events, resp. the reliability status of the network, from the set of observables. To be more precise: In this analytically tractable model for performance analysis the aspects of reliability cannot be considered explicitly.

All the model buildings presented in our paper belong to the class of investigations along the line guided by the second principle. And the way we proceeded with the respective analysis proved to be successful. The advantage of these modeling features are obviously that both aspects *Reliability and Performance behaviour* can be treated explicitly either separated or in parallel in the same model.

As we have shown by the examples in Section 7 in many cases which encompass rather general parameter structures of the systems we are able to decouple mean value analysis concerning availability and classical performance measures of queueing theory.

Some comments on the used rerouting principles may be in order here:
The STALLING-principle is a common practice, e.g. in production control: If some machine breaks down and must be repaired the production process is interrupted. Even in some big automobile firms the workers on the assembly line are able to stop the whole production process if something went wrong and / or the products are of a low-quality.
The blocking principle REPETITIVE SERVICE – RANDOM DESTINATION (RS–RD) is a well established approximation for some common communication protocols in systems with finite buffers.
The SKIPPING REGIME can be considered as a versatile rerouting protocol in networks with side constraints. In our opinion it is a promising feature in the realm of queueing network theory.

Some further concluding remarks concerning extensions and further research, the latter is partly already ongoing:

In this paper we focussed the presentation on closed and open BCMP networks and used the exponential networks for introductory purposes. All the results of the previous sections carry over with some more notational burdenings to the case of mixed networks as well.

For the nonexponential service time distributions in BCMP networks we followed the standard way of presentation as it can be found in the original paper [BCMP75]. The Coxian distributions' advantage is that using this framework leads to a Markovian system description with countable state space. But this class has the drawback that it is not dense (with respect to weak convergence) in the class of all distributions on $\mathbb{R}_+$. If necessary this drawback can be removed easily by using instead another *phase–type* distribution which is dense in the class of all distributions on $\mathbb{R}_+$, e.g., the class of finite mixtures of Erlangian distributions [Sch73].

A Necessities for the proof of Theorem 5.6.

The proof needs some properties of time-reversed Markov processes originating from Markov processes which are not necessarily reversible in time. We give a short survey according to [Kel79], Chapter 1.7.

Definition A.1 (Time-reversed process) *Let $X = (X(t) : t \in \mathbb{R})$ be a homogenous Markov process on $\mathcal{Z}$. Then the process $\bar{X} = \left(\bar{X}(t) := X(-t) : t \in \mathbb{R}\right)$ is called the time-reversed process assigned to X.*

Lemma A.2 *If the Markov process X is stationary with transition rates $q(z,z')$ for $z,z' \in \mathcal{Z}$ and equilibrium distribution $\pi(z)$, $z \in \mathcal{Z}$, then the time-reversed process $\bar{X}$ is a stationary Markov process on $\mathcal{Z}$ with transition rates*

$$\bar{q}(z,z') = \frac{\pi(z')q(z',z)}{\pi(z)}, \text{ for } z,z' \in \mathcal{Z}$$

and the same equilibrium distribution $\pi(z)$, $z \in \mathcal{Z}$.

Lemma A.3 *Let X be a stationary Markov process on $\mathcal{Z}$ with transition rates $q(z,z')$ for $z,z' \in \mathcal{Z}$. If there exists a collection of numbers $\bar{q}(z,z')$, $z,z' \in \mathcal{Z}$, such that*

$$\bar{q}(z) = q(z), \; z \in \mathcal{Z}$$

and a collection of positive numbers $\pi(z)$, $z \in \mathcal{Z}$, *summing to one such that*

$$\pi(z')q(z',z) = \pi(z)\bar{q}(z,z') \text{ for } z,z' \in \mathcal{Z}, \tag{37}$$

then $\bar{q}(z,z')$, $z,z' \in \mathcal{Z}$, *are the transition rates of the time-reversed process* $X(-t)$ *and* $\pi(z)$, $z \in \mathcal{Z}$ *is the equilibrium distribution of both processes* X *and* $\bar{X}$.

References

[AN63] B. Avi-Itzhak and P. Naor. Some queuing problems with the server station subject to breakdown. *Operations Research*, 11, pp. 303-320, 1963.

[BCMP75] F. Baskett, M. Chandy, R. Muntz and F.G. Palacios. Open, closed and mixed networks of queues with different classes of customers. *Journal of the Association for Computing Machinery*, 22, 248-260, 1975.

[BFH88] M. Bär, K. Fischer, G. Hertel. *Leistungsfähigkeit, Qualität, Zuverlässigkeit.* transpress VEB Verlag für Verkehrswesen, Berlin 1988

[BGdT98] G. Bolch, S. Greiner, H. de Meer and K. S. Trivedi. *Queueing networks and Markov chains.* John Wiley, New York, 1998.

[BNPO01] S. Balsamo, V. De Nitto Persone, R. Onvural. *Analysis of Queueing Networks with Blocking.* Kluwer academic publishers, 2001.

[CM96] R. Chakka and I. Mitrani. Approximate solutions for open networks with breakdowns and repairs. In F. P. Kelly, S. Zachary and I. Ziedins, editors, *Stochastic Networks - Theory and applications*, chapter 16, pp. 267-280, Clarendon Press, Oxford, 1996.

[Dad01] H. Daduna. Stochastic networks with product form equilibrium. In D. N. Shanbhag and C. R. Rao, editors, *Stochastic Processes: Theory and Methods*, volume 19 of Handbook of Statistics, 11, pp. 309-364, 2001.

[Dij93] N. M. Van Dijk. *Queueing Networks and Product Forms - A Systems Approach.* Wiley, Chichester, 1993.

[Dos90] B. T. Doshi. Single server queues with vacations. In H. Takagi, editor, *Stochastic Analysis of Computer and Communications Systems* pp.217 - 267, North-Holland Amsterdam 1990.

[DS96] H. Daduna and R. Szekli. A queueing theoretical proof of increasing property of Polya frequency functions. *Statistics and Probability Letters*, 26, 233-242, 1996.

[EF98] A. Economou, D. Fakinos. Product form stationary distributions for queueing networks with blocking and rerouting. *Queueing Systems and Their Applictions*, 30, pp. 251 - 260, 1998.

[FMH93] W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, 18, pp. 149-171, 1993.

[GLT01] R. German, J. Luethi, M. Telek, editors. *Proceedings of the Fifth International Workshop on Performability Modeling of Computer and Communication Systems,* September 15-16, 2001, Erlangen.

[GN67] W. J. Gordon, G. F. Newell. Closed queueing systems with exponential servers. *Operations Research*, 15, pp. 254-265, 1967.

[GP87] E. Gelenbe, G. Pujolle. *Introduction to queueing networks.* Wiley, New York, 1987.

[HMRT01] B. R. Haverkort, R. Marie, G. Rubino and K. Trivedi, editors, *Performability Modelling - Techniques and Tools.* Wiley, 2001.

[Jac57] J. R. Jackson. Networks of waiting lines. *Operations Research*, 5, pp. 518-521, 1957.

[Jac63] J. R. Jackson. Jobshop-like queueing systems. *Management Science*, 10, pp. 131-142, 1963.

[Kei62] J. Keilson. Queues subject to service interruption. *The Annals of Mathematical Statistics*, 33 No. 4, pp. 1314-1322, 1962.

[Kel79] F. P. Kelly. *Reversibility and stochastic networks.* Wiley, New York, 1979.

[LAP01] H. W. Lee, B. Y. Ahn, N. I. Park. Decompositions of the queue length distributions in the MAP/G/1 queue under multiple and single vacations with N-policy. *Stochastic Models*, 17, pp. 157 - 190, 2001.

[MAI68] I. Mitrani and B. Avi-Itzhak. A many-server queue with service interruptions. *Operations Research*, 16, pp. 628-638, 1968.

[MCl88] B. Müller-Clostermann. An approximate product form for a class of degradable queueing networks. *Performance Evaluation*, 8, pp. 165-172, 1988.

[Mel82] B. Melamed. Sojourn times in queueing networks. *Mathematics of Operations Research*, 7, pp. 223-244, 1982.

[Mey84] J.F. Meyer. Performability modeling of distributed real-time systems. In G. Iazeolla, P.J. Courtois and A. Hordijk, editors, *Mathematical Computer Performance and Reliability*, pp. 361-372, North-Holland, Amsterdam, 1984.

[Mik88] N. Mikou. A two-node Jackson's network subject to breakdowns. *Stochastic Models*, 4, pp. 523-552, 1988.

[MIKS95] N. Mikou, O. Idrissi-Kacimi and S. Saadi. Two processes interacting only during breakdown: the case where the load is not lost. *Queueing Systems*, 19, pp. 301-317, 1995.

[Mit74] I. Mitrani. Networks of unreliable computers. In E. Gelenbe and R. Mahl, editors, *Computer Architectures and Networks*, pp. 359-374, North-Holland, Amsterdam, 1974.

[MO67] A.W. Marshall, I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 2, pp. 30-44, 1967.

[MW94] I. Mitrani and P. E. Wright. Routing in the presence of breakdowns. *Performance Evaluation*, 20, pp. 151-164, 1994.

[Onv90] R. O. Onvural. Closed queueing networks with blocking. In H. Takagi, editor, *Stochastic Analysis of Computer and Communication Systems*, pp. 499-528, North-Holland, Amsterdam, 1990.

[Per90] H. G. Perros. Approximation algorithms for open queueing networks with blocking. In H. Takagi, editor, *Stochastic Analysis of Computer and Communication Systems*, pp. 451-498, North-Holland, Amsterdam, 1990.

[Sch73] R. Schassberger. *Warteschlangen.* Springer, Wien, 1973.

[Sch84] R. Schassberger. Decomposable stochastic networks: Some observations. In F. Baccelli and G. Fayolle, editors, *Modelling and Performance Evaluation Methodology*, volume 60 of *Lecture Notes in Control and Information Sciences*, chapter IV, pp. 137-150, Springer, Berlin, 1984. Proceedings of the International Seminar, Paris, France, January 24-26, 1983.

[Ser93] R. F. Serfozo. Queueing networks with dependent nodes and concurrent movements. *Queueing Systems*, 13, pp. 143-182, 1993.

[Ser99] R. F. Serfozo. *Introduction to Stochastic Networks.* Applications of Mathematics 44, Springer, New York, 1999.

[Thi63] K. Thiruvengadam. Queuing with breakdowns. *Operations Research*, 11, pp. 62-71, 1963.

[WC58] H. White and L. S. Christie. Queuing with preemptive priorities or with breakdown. *Operations Research*, 6, pp. 79-95, 1958.