

PRODUCT FORMS FOR AVAILABILITY MODELS

ERIC SMEITINK

PTT Research, St Paulusstraat 4, 2264 XZ Leidschendam, The Netherlands

NICO M. VAN DIJK

Faculteit Econometrie, Universiteit van Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands

AND

BOUDEWIJN R. HAVERKORT

Faculteit der Informatica, Universiteit Twente, Postbus 217, 7500 AE Enschede, The Netherlands

SUMMARY

This paper shows and illustrates that product form expressions for the steady state distribution, as known for queueing networks, can also be extended to a class of availability models. This class allows breakdown and repair rates from one component to depend on the status of other components. Common resource capacities and repair priorities, for example, are included. Conditions for the models to have a product form are stated explicitly. This product form is shown to be insensitive to the distributions of the underlying random variables, i.e. to depend only on their means. Further it is briefly indicated how queueing for repair can be incorporated. Novel product form examples are presented of a simple series/parallel configuration, a fault tolerant database system and a multi-stage interconnection network.

KEY WORDS Availability modelling Product form Insensitivity

1. INTRODUCTION

The measures commonly used to quantify the functioning of a system are reliability and availability. The reliability function is typically of interest for systems that must remain operational during their whole (limited) mission length, such as air or spacecraft systems. If components can be repaired and if some down time can be tolerated, then availability is the more appropriate concept. There are several availability measures. An overview of numerical methods for calculating transient measures such as point- and interval-availability is given in Reference 1.

In this paper we focus on the steady state availability of repairable computer systems. We consider a system consisting of N components that are alternatively up and down. The simplest case occurs when all components have independent up times and down times with mean λ_h^{-1} respectively μ_h^{-1} for component h . In this case note that $p_h = \mu_h / (\mu_h + \lambda_h)$ corresponds to the marginal steady state probability that component h is up, so that the steady state probability $\pi(H)$ that the components in the set H are down while the others are up is easily concluded to be given by

$$\pi(H) = \prod_{h \notin H} p_h \prod_{h \in H} (1 - p_h) \quad (1.1)$$

The asymptotic properties of this model, which is often used in reliability studies of (large) communication networks, are extensively studied in, for example, Reference 2. A typical measure of interest in this context is the steady state probability that two nodes can communicate via a network of links that are subject to failure. This probability can in principle be calculated from the steady state distribution π (cf. References 3, 4, 5, 6).

Fault tolerant computer systems often exhibit *state dependencies*, by which the breakdown or repair rate of one component can be accelerated or slowed down depending on the up or down status of other (neighbouring) components. For example, if all components share a common resource for repair, the down times will most likely increase with an increasing number of components that are down. Under exponential assumptions, such systems can be analysed using continuous time Markov models. In that case, the steady state distribution can in principle be obtained by numerically solving a system of linear equations. This can be computationally expensive, however, and if the cardinality of the state space is large special techniques have to be used, such as sparse matrix storage and successive overrelaxation (cf. Reference 7). In Reference 8 a bounding technique based on state space aggregation is presented for obtaining the steady state availability of highly reliable systems.

The primary purpose of this paper is to illustrate that product form expressions for steady state distributions, which have been extensively studied for queueing networks, also apply to a class of availability models. This class is characterized by imposing conditions on the state dependent breakdown and repair rates. Although from a mathematical point of view the product form expression is not new, as it is based on the notion of reversibility (cf. References 9 and 10), its recognition for the present purpose appears to be new. In particular, as state dependent breakdown and repair rates arise naturally, the interpretation of the actual transition rates is different. If in some configuration H a certain transition is prohibited, for example the breakdown of component $h \notin H$, the corresponding transition rate is zero, i.e. the repair of that component is stopped. This situation occurs, for example, in the model of a series system considered in Reference 11 (pp. 194–201). As the system is a pure series configuration, system failure coincides with component failure and it is therefore natural to assume that, while a failed component undergoes repair, all other components remain in 'suspended operation'. When the repair of the failed component is complete, the remaining components resume operation. At that instant they are not 'as good as new', but rather as good as they were when the system stopped operating. The steady state results obtained in Reference 11 turn out to be insensitive, i.e. to depend only on the means of the underlying random variables.

By introducing speed functions, i.e. state dependent breakdown and repair rates, we can generalize the above model of a series system and consider arbitrary systems. We then state sufficient conditions on the speed functions that guarantee a product form steady state distribution. These so-called product form conditions impose a special structure on the stochastic process of components going up and down. Loosely put, we require that in the steady state the system exhibits a certain *local balance* property. Local balance properties are closely connected to insensitivity.^{12,13,14}

This paper is organized as follows. In Section 2 we formulate the model and illustrate how the speed functions can be used to model failure and repair dependencies between the components. In Section 3 we first formulate the product form conditions. Then, under exponential assumptions, we show how these conditions naturally arise when requiring local balance and obtain the steady state probabilities $\pi(H)$ that the system is in configuration H . In Section 4 we consider a slight extension of our original model and illustrate, again under exponential assumptions, how queueing for repair can be incorporated. In Section 5 we show

that the steady state probabilities $\pi(H)$ are insensitive. In Section 6 we give three examples to illustrate how various systems can be modelled and how the product form conditions can be verified. Section 7 concludes the paper.

2. MODEL

We consider a system consisting of N components where each component is alternatively up (being able to work) and down (not able to work and needing repair). The set of all components is denoted by C and a configuration $H \subset C$ denotes that the components in the set H are down while the other components are up. The configuration in which all components are up (none is down) is denoted by \emptyset .

At time $t = 0$ all components start their first up period. During its i th up period, component h performs a random amount of work $W_{h,i}$ before going down. When component h goes down for the i th time it requires a random amount of repair $R_{h,i}$. Upon completion of its repair a component immediately returns to the up mode. We assume that all these random variables $\{W_{h,i}\}_{i \geq 1}$ and $\{R_{h,i}\}_{i \geq 1}$ are mutually independent and that, for each fixed $h \in C$, the random variables $\{W_{h,i}\}_{i \geq 1}$ respectively $\{R_{h,i}\}_{i \geq 1}$ are identically distributed. Consequently, $\{W_{h,i}\}_{i \geq 1}$ and $\{R_{h,i}\}_{i \geq 1}$ are renewal sequences. We denote the distribution functions of the generic variables W_h and R_h by F_h , respectively G_h .

The speed at which a component $h \in H$ is repaired is denoted by $\delta(h | H)$, i.e. the remaining amount of repair that component h requires before going up again is provided at the configuration dependent rate $\delta(h | H)$. Likewise, the speed at which a component $h \notin H$ works when the system is in configuration H is given by $\beta(h | H)$. We will refer to β and δ as the speed functions. Further, if we write $\beta(h | H)$ ($\delta(h | H)$) we implicitly assume that $h \notin H$ ($h \in H$).

We denote by $\mathcal{H} \subset 2^C$ the set of configurations, including \emptyset , which can be reached from \emptyset by components going up and down (2^C denotes the power set of C). For convenience, we will denote by $H + h$ the configuration in which, apart from the components in H , an additional component $h \notin H$ is down. Similarly, $H - h$ denotes the configuration in which the components in H are down except for component $h \in H$.

Apart from the renewal assumptions above we will require the following *regularity conditions* for all components $h \in C$.

R1 The speed functions $\beta(h | \cdot)$ and $\delta(h | \cdot)$ are bounded.

R2 $0 < E[W_h], E[R_h] < \infty$

R3 The distribution functions F_h and G_h are continuously differentiable. We denote the probability densities by f_h , respectively g_h . Furthermore, these densities are bounded.

Note that **R2** directly implies **R3** if F_h and G_h are exponentially distributed. In this case we also have that **R1** and **R2** together imply that

$$H \in \mathcal{H} \text{ and } \beta(h | H) > 0 \Rightarrow H + h \in \mathcal{H} \tag{2.1}$$

and

$$H \in \mathcal{H} \text{ and } \delta(h | H) > 0 \Rightarrow H - h \in \mathcal{H} \tag{2.2}$$

Condition **R3** ensures that (2.1) and (2.2) also hold in the general case, i.e. that the set of configurations \mathcal{H} that can be reached in the general case is the same as the set of configurations that can be reached under exponential assumptions.

Interpretation of the speed functions

The actual *time* that component h remains up depends on the random amount of work, W_h , that it performs before going down and on the varying speed(s) $\beta(h | \cdot)$ with which it works, i.e. on the configurations that the system visits during the up time of component h . In the same way the time that a component h remains down depends on the amount of repair, R_h , that it needs and on the speed(s) $\delta(h | \cdot)$ with which it is repaired. Below we give some examples of how the functions β and δ can be used as delay or acceleration factors to model failure and repair dependencies between the components.

- (a) $\beta(h | H) = 0$ reflects that component h , although it is up, does not work (cannot be used) if the components in the set H are down. Consequently, component h cannot go down if the system is in configuration H . Note that our model is not applicable to circumstances where the 'ageing process' continues while a component is unused.
- (b) $\beta(h | H + h') > \beta(h | H)$ indicates that component h works faster if the additional component h' has failed. This can be used to model situations where work is shared between components that are up.
- (c) $\delta(h | H + h + h') < \delta(h | H + h)$ indicates that the repair of component h is slowed down if an additional component h' has failed. This situation occurs when components share a common repair facility.
- (d) $\delta(h | \{h, h'\}) > \delta(h' | \{h, h'\}) = 0$ models the (pre-emptive) priority of the repair of component h over that of component h' .

Steady state behaviour

We are interested in the steady state behaviour of the stochastic process $\tilde{X} = \{\tilde{X}(t), t \geq 0\}$ on \mathcal{H} , defined by letting $\tilde{X}(t) = H$ if at time t the configuration of the system is H . It will be clear that, in general, the process \tilde{X} is not a Markov process, as the future evolution of the system depends on the residual amounts of work and repair. In order to imbed the process \tilde{X} in a Markov process X we define for a given configuration H the vector $s = (s_1, \dots, s_N) \in \mathcal{S} = \mathbf{R}_+^N$ by

$$s_h = \begin{cases} \text{the residual amount of work for component } h, & \text{if } h \notin H \\ \text{the residual amount of repair for component } h, & \text{if } h \in H \end{cases}$$

Owing to the independence assumptions on the random variables $\{R_{h,i}\}_{i \geq 1}$ and $\{W_{h,i}\}_{i \geq 1}$ the state of the system is then completely determined by a pair $(H, s) \in \mathcal{H} \times \mathcal{S}$ and, defining $X(t) = (H, s)$ if at time $t \geq 0$ the system is in state (H, s) , it is easily verified that the stochastic process $X = \{X(t), t \geq 0\}$ on $\mathcal{H} \times \mathcal{S}$ has the desired Markov property. The regularity conditions **R1**–**3** ensure that X has a limiting probability density function p (see, for example, Reference 15), which we will obtain in Section 5. For all $H \in \mathcal{H}$ the steady state probability $\pi(H)$ that the system is in configuration H is then given by

$$\pi(H) = \lim_{t \rightarrow \infty} P(\tilde{X}(t) = H) = \lim_{t \rightarrow \infty} P(X(t) \in (H, \mathcal{S})) = \int_{\mathcal{S}} p(H, s) ds \quad (2.3)$$

3. PRODUCT FORM CONDITIONS

It will be convenient to define for an arbitrary state $H = \{h_1, \dots, h_n\} \in \mathcal{H}$ ($H \neq \emptyset$) and bijection $\gamma: \{1, \dots, n\} \rightarrow \{h_1, \dots, h_n\}$ the sets H_k^γ for $k \geq 0$ by $H_0^\gamma = \emptyset$ and

$$H_k^\gamma = \{\gamma(1), \dots, \gamma(k)\}, \quad k = 1, \dots, n$$

The bijection γ indicates a possible ordering of the components of H , which might indicate, for example, that component $\gamma(i)$ went down before component $\gamma(j)$ if $i < j$.

Definition 3.1

An ordering γ of $H = \{h_1, \dots, h_n\}$ is called a *breakdown path* if

$$\beta(\gamma(k) | H_{k-1}^\gamma) > 0, \quad k = 1, \dots, n$$

and a *repair path* if

$$\delta(\gamma(k) | H_k^\gamma) > 0, \quad k = 1, \dots, n$$

The set of all possible breakdown paths for state H is denoted by $\Gamma_\beta(H)$ and the set of all repair paths is denoted by $\Gamma_\delta(H)$.

Note that $H \in \mathcal{H}$ does not necessarily imply that $H_k^\gamma \in \mathcal{H}$ and that in principle \mathcal{H} may contain states for which no breakdown path and/or repair path exists. However, this pathological case will be excluded by the conditions **PFC 1** and **PFC 2** below. The product form conditions are the following.

PFC 1

For any state $H \in \mathcal{H} (H \neq \emptyset)$ we must have that

$$\sum_{h \in H} \delta(h | H) > 0$$

i.e at least one component is being repaired and as a result $\Gamma_\delta(H) \neq \emptyset$.

PFC 2

For any state $H \in \mathcal{H} (H \neq \emptyset)$ we must have that

$$\delta(h | H) = 0 \Leftrightarrow \beta(h | H - h) = 0 \text{ for all } h \in H \text{ such that } H - h \in \mathcal{H}$$

i.e. if component h cannot be repaired in configuration H it cannot break down in configuration $H - h$, and vice versa. As a result we have that $\Gamma_\beta(H) = \Gamma_\delta(H)$.

PFC 3

For any state $H = \{h_1, \dots, h_n\} \in \mathcal{H} (H \neq \emptyset)$ and repair path $\gamma \in \Gamma_\delta(H)$ we must have that

$$K(H) = \prod_{k=1}^n \frac{\beta(\gamma(k) | H_{k-1}^\gamma)}{\delta(\gamma(k) | H_k^\gamma)} \tag{3.1}$$

where $K(H)$ is some strictly positive value depending on H only. We define $K(\emptyset) = 1$. This key condition is related to the well-known Kolmogorov criterion for a Markov chain to be reversible (cf. Reference 10, pp. 21–25]). A further discussion of this condition and an interpretation are given in Section 3.2.

The speed functions β and δ often suggest a form for the function $K(\cdot)$, in which case **PFC 3** can be checked using the equivalent condition **PFC 3'** stated below.

PFC 3'

For all states $h \in \mathcal{H}$ ($H \neq \emptyset$) and all components $h \in H$ such that $\delta(h | H) > 0$ we have that

$$K(H) = K(H - h) \frac{\beta(h | H - h)}{\delta(h | H)} \quad (3.2)$$

Remark 3.2

As we will frequently use **PFC 3'** let us briefly show that the two are indeed equivalent. First observe that for $H = \{h_1, \dots, h_n\} \in \mathcal{H}$ ($H \neq \emptyset$) and $h \in H$ such that $\delta(h | H) > 0$ there exists a repair path $\gamma \in \Gamma_\delta(H)$ with $\gamma(n) = h$, i.e. the repair of component h is completed first. Using this path in (3.1) we obtain

$$K(H) = \left(\prod_{k=1}^{n-1} \frac{\beta(\gamma(k) | H_{k-1}^\gamma)}{\delta(\gamma(k) | H_k^\gamma)} \right) \frac{\beta(\gamma(n) | H_n^\gamma)}{\delta(\gamma(n) | H_n^\gamma)} = K(H - h) \frac{\beta(h | H - h)}{\delta(h | H)}$$

Thus **PFC 3** implies **PFC 3'**. Repeatedly using **PFC 3'** in (3.2) yields that

$$K(H) = K(\emptyset) \prod_{k=1}^n \frac{\beta(\gamma(k) | H_{k-1}^\gamma)}{\delta(\gamma(k) | H_k^\gamma)}$$

for each repair path $\gamma \in \Gamma_\delta(H)$. Noting that we defined $K(\emptyset) = 1$ it follows that **PFC 3'** implies **PFC 3**.

Source balance

In order to illustrate the role of the conditions **PFC 1–3** let us first assume that W_h and R_h are exponentially distributed with mean $E[W_h] = \lambda_h^{-1}$ and $E[R_h] = \mu_h^{-1}$ respectively. Owing to the memoryless property of the exponential distribution, the process $\tilde{X} = \{\tilde{X}(t), t \geq 0\}$ defined in Section 2 is a Markov process at \mathcal{H} , for which a unique steady state distribution π exists if \tilde{X} is irreducible. This is guaranteed by condition **PFC 1**, as the existence of a repair path for each state $H \in \mathcal{H}$ guarantees that all states communicate with the state \emptyset .

The steady state distribution π can be found as the unique solution (up to normalization) of the *global balance* equations, which require that for every state $H \in \mathcal{H}$ the ‘total rate of change’ is equal to zero, i.e.

$$\sum_{\{h \notin H: H+h \in \mathcal{H}\}} [\pi(H)\beta(h | H)\lambda_h - \pi(H+h)\delta(h | H+h)\mu_h] + \sum_{\{h \in H: H-h \in \mathcal{H}\}} [\pi(H)\delta(h | H)\mu_h - \pi(H-h)\beta(h | H-h)\lambda_h] = 0 \quad (3.3)$$

where we used the observation that $\delta(h | H) > 0 \Rightarrow H - h \in \mathcal{H}$ and $\beta(h | H) > 0 \Rightarrow H + h \in \mathcal{H}$. In order to find a closed-form solution to the global balance equations we require the stronger *source balance* property to hold, which in our model requires that for every state $H \in \mathcal{H}$:

$$\begin{aligned} & \text{the probability flow into state } H \text{ due to the breakdown of component } h \text{ equals} \\ & \text{the probability flow out of that same state } H \text{ due to the repair of component } h, \quad (3.4) \\ & \text{and vice versa} \end{aligned}$$

This notion of source balance is related to job–local balance as introduced in Reference 12 and local balance as introduced in Reference 13, which are known to be responsible for insensitive

product form results. It is readily verified that the global balance equations (3.3) are satisfied if the source balance property (3.4) holds, i.e. if we can find a probability distribution π such that for every state $H \in \mathcal{H}, H \neq \emptyset$

$$\pi(H) \delta(h | H)_{\mu_h} = \pi(H-h) \beta(h | H-h)_{\lambda_h} \tag{3.5}$$

for every component $h \in H$ such that $H-h \in \mathcal{H}$. In Theorem 3.3 below we will see that this is possible provided conditions **PFC 1–3** are satisfied. In Section 5 we will show that the same result holds true for generally distributed quantities of work and repair, i.e. that the steady state probabilities $\pi(H)$ given in (3.6) below are insensitive.

Theorem 3.3

Provided that **PFC 1** holds, the conditions **PFC 2** and **PFC 3** are both necessary and sufficient for source balance. Further, if the conditions **PFC 1–3** are satisfied, the steady state distribution π of \tilde{X} is given by

$$\pi(H) = cK(H) \prod_{h \in H} \frac{\lambda_h}{\mu_h} \tag{3.6}$$

with c a normalizing constant and $K(H)$ given by (3.1).

Proof

That **PFC 2** is a necessary condition for source balance follows directly from the irreducibility of \tilde{X} and the regularity conditions **R1–2**, which imply $\pi(H) > 0$ for all $H \in \mathcal{H}$. Let $H = \{h_1, \dots, h_n\}$ ($H \neq \emptyset$) and let $\gamma \in \Gamma_\delta(H)$ denote an arbitrary repair path for state H . By repeatedly using (3.5) it readily follows that source balance implies

$$\pi(H) = \pi(\emptyset) \left(\prod_{k=1}^n \frac{\beta(\gamma(k) | H_{k-1}^\gamma)}{\delta(\gamma(k) | H_k^\gamma)} \right) \left(\prod_{k=1}^n \frac{\lambda_{h_k}}{\mu_{h_k}} \right) \tag{3.7}$$

As the ratio between $\pi(H)$ and $\pi(\emptyset)$ must be constant, we conclude that **PFC 3** is also necessary for source balance.

All that remains to verify is that π given by (3.6) satisfies the source balance equations (3.5), as these imply that each separate term between brackets in (3.3) equals zero. Owing to **PFC 2** we need only consider states $H \in \mathcal{H}$ and components $h \in H$ with $\delta(h | H) > 0$. As for any such state we also have $H-h \in \mathcal{H}$ with $\beta(h | H-h) > 0$ we obtain from (3.6) and **PFC 3'** that

$$\frac{\pi(H)}{\pi(H-h)} = \frac{K(H)\lambda_h}{K(H-h)\mu_h} = \frac{\beta(h | H-h)\lambda_h}{\delta(h | H)\mu_h}$$

which completes the proof. □

Remark 3.4

The interpretation of **PFC 3** is that given the starting point $\emptyset \in \mathcal{H}$ any path in the state space which ultimately returns to \emptyset must have the same probability whether this path is traced in one direction or the other.

Remark 3.5

The equations (3.5) are also called *detailed balance* equations,¹⁴ a term reserved to express balance between any pair of states, in this example that is between states H and $H-h$. When

applied to the same state description, detailed balance implies source balance, but in Section 4.2 (Equation (4.4)) we will see that source balance does not necessarily imply detailed balance

4. QUEUEING FOR REPAIR

The model described in Section 2 above is primarily suited for systems with a hierarchic structure. As the breakdown and repair speeds depend only on the set of components that are down and not on the order in which they broke down, product form results for systems with an FIFO queue of failed components in front of a limited number of repair facilities, for example, cannot be directly concluded using this model.

4.1. Identical components

The simplest such case is the following basic machine interference model (see Reference 10, p. 100), where all components $h \in C$ are assumed to be identical and where W_h and R_h are exponentially distributed, each with the same mean $E[W_h] = \lambda^{-1}$ respectively $E[R_h] = \mu^{-1}$. The state of the system is now completely determined by the number of components that are down. We denote by $\phi(n)$, $n = 0, \dots, N-1$, and $\psi(n)$, $n = 1, \dots, N$, the total work, respectively repair speed, when n components are down. For example, $\psi(n) = 1$ models an FIFO queue for repair, whereas $\psi(n) = n$ models independent repairs. The situation that only one component is actually working with the others in cold standby (not working) can be modelled by letting $\phi(n) = 1$, reflecting an 'FIFO queue for breakdowns'.

We define n_{\max} as the maximum number of components that can be down at the same time. As initially all components are up $n_{\max} = \min \{n: \phi(n) = 0\}$. We assume that ϕ and ψ are bounded and that $\psi(n) > 0$, $n = 1, \dots, n_{\max}$. These conditions are the equivalent of **R 1** and **PFC 1**. Then, denoting by $\bar{X}(t) = n$ the event that n components are down at time $t \geq 0$, it can easily be verified that the steady state distribution $\bar{\pi}$ of the Markov chain $\bar{X} = \{\bar{X}(t), t \geq 0\}$ (a birth-death process) is given by

$$\bar{\pi}(n) = c \left(\prod_{k=1}^n \frac{\phi(k-1)}{\psi(k)} \right) \left(\frac{\lambda}{\mu} \right)^n, \quad n = 0, \dots, n_{\max} \quad (4.1)$$

with c a normalizing constant. Thus \bar{X} has a product form steady state distribution, even though conditions **PFC 2** and **PFC 3** are not necessarily satisfied. Instead, however, we have assumed that all components are identical with exponentially distributed quantities of work W_h and repair R_h .

Remark 4.1

The steady state distribution $\bar{\pi}$ satisfies the detailed balance equations (3.5), if we only specify the total number of components, n , that are down. This seems to conflict with Remark 3.5, since the source balance property (3.4) does not generally hold. Note, however, that here we work with a different, aggregated, state description.

4.2. Non-identical components

In relation with queueing theory^{10,16} the conditions of Section 4.1 may be relaxed. Let us illustrate this by considering one concrete situation in which the components of some subset

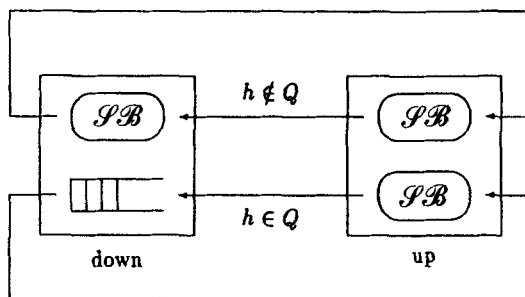


Figure 1. Queueing network interpretation

$Q \subset C$ join an FIFO queue for repair. In order to provide insight we will constructively investigate the consequences of allowing queueing in this case. These consequences are summarized in Theorem 4.2 below, which states that a product form steady state distribution is still possible, provided the additional conditions **Q 1–3** are satisfied.

It will be convenient to identify each component with a job, circulating between node ‘up’ and node ‘down’, in a two-node queueing network (see Reference 7 and Reference 10, pp. 99–105). We have depicted the network corresponding to the model with one FIFO queue for repair in Figure 1. Jobs $h \notin Q$ can, for example, be served in a processor sharing manner, receiving their required amount of repair, R_h , at speed $\delta(h|H)$ which decreases with an increasing number of jobs in node ‘down’. Similarly, accelerated ageing processes in the case of many failed components can be modelled by assigning a separate server to each job at node ‘up’, consuming the amount of work W_h at speed $\beta(h|H)$ which increases with an increasing number of jobs in node ‘down’. The parts of the network where the source balance property (3.4), with respect to the augmented state description (H, ξ) (see below), will be seen to hold are indicated by PB .

In order to model the positions in the queue we augment our description of a configuration H with a bijection $\xi: \{1, \dots, m\} \rightarrow H \cap Q$, where $m = |H \cap Q|$ and $\xi(i)$ denotes the component that occupies position i in the queue, $i = 1, \dots, m$. This configuration is denoted by (H, ξ) . For notational convenience we will write ξ_i rather than $\xi(i)$ and omit parentheses if there is no danger of ambiguity. The configuration with one additional component $h \in Q$ down and in the last position of the queue (arrival) is denoted by $(H + h, (\xi_{1 \rightarrow m}, h))$. Similarly, the configuration with the component ξ_1 up instead of down (departure) is denoted by $(H - \xi_1, (\xi_{2 \rightarrow m}))$, where we use the convention that $\xi_{i \rightarrow j} = \emptyset$ for $i > j$. The configuration-dependent breakdown and repair speeds are denoted $\bar{\beta}(h|H, \xi)$ and $\bar{\delta}(h|H, \xi)$, where queueing for repair of course implies that for each component $h \in H \cap Q$ the repair speed $\bar{\delta}(h|H, \xi) = 0$ if $h \neq \xi_1$, since only the component that occupies the first position in the queue is being repaired. We assume that the speed functions $\bar{\beta}$ and $\bar{\delta}$ are bounded and that in each state (H, ξ) with $H \neq \emptyset$ there is at least one component that is being repaired with a positive speed. These are the equivalent of **R 1** and **PFC 1**. Further we assume that W_h and R_h are exponentially distributed with mean $E[W_h] = \lambda_h^{-1}$ and $E[R_h] = \mu_h^{-1}$, so that we can define the Markov process $\bar{X} = \{\bar{X}(t), t \geq 0\}$ by letting $\bar{X}(t) = (H, \xi)$ if at time t the system is in state (H, ξ) .

Balance equations

Analogous to the case without queueing, the unique steady state distribution $\bar{\pi}$ for \bar{X} can be found as the normalized solution of the global balance equations, which require that for

all states (H, ξ)

$$\begin{aligned}
& \sum_{h \in H^c \cap Q^c} [\bar{\pi}(H, \xi) \bar{\beta}(h | H, \xi) \lambda_h - \bar{\pi}(H + h, \xi) \bar{\delta}(h | H + h, \xi) \mu_h] \\
& + \sum_{h \in H \cap Q^c} [\bar{\pi}(H, \xi) \bar{\delta}(h | H, \xi) \mu_h - \bar{\pi}(H - h, \xi) \bar{\beta}(h | H - h, \xi) \lambda_h] \\
& + \sum_{h \in H^c \cap Q} [\bar{\pi}(H, \xi) \bar{\beta}(h | H, \xi) \lambda_h - \bar{\pi}(H + h, (h, \xi_1 \rightarrow m)) \bar{\delta}(h | H + h, (h, \xi_1 \rightarrow m)) \mu_h] \\
& + [\bar{\pi}(H, \xi) \bar{\delta}(\xi_1 | H, \xi) \mu_{\xi_1} - \bar{\pi}(H - \xi_m, (\xi_1 \rightarrow m - 1)) \bar{\beta}(\xi_m | H - \xi_m, (\xi_1 \rightarrow m - 1)) \lambda_{\xi_m}] = 0 \quad (4.2)
\end{aligned}$$

It will be clear that source balance cannot be satisfied at the queue, corresponding to the last term of (4.2): the state (H, ξ) can be entered due to the breakdown of component ξ_m , but since ξ_m is the last component in the queue, state (H, ξ) can in general not be left due to the repair of component ξ_m (the only exception being the case $m = 1$).

However, we can still try to solve the global balance equations in a local manner, i.e. by requiring that each separate term between brackets in (4.2) above equals zero. As in the case without queueing, the terms corresponding to components $h \notin Q$ are equal to zero, provided that for all components $h \in H \cap Q^c$

$$\bar{\pi}(H, \xi) \bar{\delta}(h | H, \xi) \mu_h = \bar{\pi}(H - h, \xi) \bar{\beta}(h | H - h, \xi) \lambda_h \quad (4.3)$$

reflecting source balance for all components $h \notin Q$, both in node ‘up’ and node ‘down’ (see Figure 1). By considering the rate of change in state $(H - \xi_1, (\xi_2 \rightarrow m))$ rather than (H, ξ) , we see that the terms in the third sum all equal zero, provided that

$$\bar{\pi}(H, \xi) \bar{\delta}(\xi_1 | H, \xi) \mu_{\xi_1} = \bar{\pi}(H - \xi_1, (\xi_2 \rightarrow m)) \bar{\beta}(\xi_1 | H - \xi_1, (\xi_2 \rightarrow m)) \lambda_{\xi_1} \quad (4.4)$$

This equation reflects that the probability flow into state $(H - \xi_1, (\xi_2 \rightarrow m))$ due to the repair of component ξ_1 must equal the probability flow out of that same state due to the breakdown of component ξ_1 , corresponding to source balance for components $h \in Q$ in node ‘up’. Note that (4.4) is not a detailed balance equation (see Remark 3.5). The last equation is just the last term of (4.2) corresponding to the queue (no source balance),

$$\bar{\pi}(H, \xi) \bar{\delta}(\xi_1 | H, \xi) \mu_{\xi_1} = \bar{\pi}(H - \xi_m, (\xi_1 \rightarrow m - 1)) \bar{\beta}(\xi_m | H - \xi_m, (\xi_1 \rightarrow m - 1)) \lambda_{\xi_m} \quad (4.5)$$

Conditions and result

As in Section 3 (see Equation (3.7)) we can repeatedly use the equations (4.3)–(4.5) to obtain the (fixed) ratio between the steady state probabilities $\bar{\pi}(H, \xi)$ and $\bar{\pi}(\emptyset, \emptyset)$. However, if $\bar{\delta}(\xi_1 | H, \xi) > 0$ we can either use (4.4) or (4.5) to proceed with. If both $\bar{\delta}(\xi_1 | H, \xi) > 0$ and $\bar{\delta}(h | H, \xi) > 0$ for some $h \notin Q$ we may even use any of the three equations. These observations lead to the conjecture that the global balance equations (4.2) can only be solved in this local manner if the conditions Q 1–3 below are satisfied.

Q 1 All components $h \in Q$ require an exponentially distributed amount of repair R with the same mean $1/\mu$.

Q 2 The speed functions $\bar{\beta}$ and $\bar{\delta}$ are of the form

$$\bar{\beta}(h | H, \xi) = \beta(h | H) \quad \text{and} \quad \bar{\delta}(h | H, \xi) = \begin{cases} \delta(h | H), & \text{if } h = \xi_1 \text{ or } h \in Q^c \\ 0, & \text{otherwise} \end{cases} \quad (4.6)$$

i.e. apart from modelling the queue, the order ξ does not influence the speeds.

Q 3 For all states (H, ξ) and for all components $h, h' \in Q$ we must have that

$$\delta(h | H) = \delta(h' | H) \tag{4.7}$$

i.e. all components $h \in Q$ are repaired in exactly the same manner.

With these conditions we are almost back in the setting of our original model. Although β and δ are not the real speed functions in the model with queueing, we can still use the notion of (now fictitious) breakdown and repair paths as introduced in Definition 3.1 and formulate the conditions **PFC 1–3**. We then have the following result.

Theorem 4.2

If condition **Q 1** holds and the speed functions $\bar{\beta}$ and $\bar{\delta}$ are given by (4.6) for some functions β and δ that satisfy **PFC 1–3** and **Q 3**, the steady state distribution $\bar{\pi}$ of \bar{X} is given by

$$\bar{\pi}(H, \xi) = cK(H) \left(\prod_{h \in H \cap Q^c} \frac{\lambda_h}{\mu_h} \right) \left(\prod_{h \in H \cap Q} \frac{\lambda_h}{\mu} \right) \tag{4.8}$$

with c a normalizing constant and $K(H)$ given by (3.1).

Proof

Using (4.6) and condition **PFC 3'** the source balance equations (4.3) and (4.4) can be directly verified in the same manner as was done in the proof of Theorem 3.3. Equation (4.5) can be verified by first using (4.6) and the conditions **Q1** and **Q2** to write $\bar{\delta}(\xi_i | H, \xi)_{\mu_{\xi_i}} = \delta(\xi_i | H)_{\mu_{\xi_i}} = \delta(\xi_m | H)_{\mu}$. □

Remark 4.3

We have already hinted at the connection between insensitivity and source balance as related to job-local balance.¹² Without proof we state that the steady state probabilities $\bar{\pi}(H, \xi)$ depend on the distributions of W_h for $h \in C$ and R_h for $h \notin Q$ only through their means. These distributions correspond to the parts of the network in Figure 1 where source balance is satisfied, indicated by \mathcal{SB} . This partial insensitivity can be proven along the same lines as will be done for the case without queueing in Section 5.

Remark 4.4

The above extension of our model to incorporate queueing reflects only one of the many possibilities. However, we considered FIFO queueing for repair as it seems to be one of the most natural cases. Further, Theorem 4.2 can easily be extended to model several different queues. Note that components in cold standby can also be modelled by an ‘FIFO queue for breakdowns’.

5. INSENSITIVITY

In Section 4 we extended our model to incorporate queueing for repair, and saw that in order to obtain a product form steady state distribution exponential assumptions are needed. In this section we consider the model as stated in Section 2 and concentrate on insensitivity. The main

results is the technical Theorem 5.2, yielding a product form limiting probability density function $p: \mathcal{H} \times \mathcal{S} \rightarrow \mathbf{R}_+$ for the Markov process X defined in Section 2. The more practical insensitive product form expression for the steady state probabilities $\pi(H)$, $h \in \mathcal{H}$, given in Corollary 5.3, is an immediate consequence.

Although this insensitivity result can in principle be concluded from Reference 13, we prefer to give a direct proof, both for completeness and because it clearly illustrates the role of the regularity- and product form conditions. Extensions of Theorem 5.2 and Corollary 5.3 so as to include queueing for repair under appropriate conditions can be obtained by combining the results from this section and Section 4.

Remark 5.1

In the proof of Theorem 5.2 we will assume that a limiting probability density function p of X exists. That this is indeed the case can be argued under the sufficient regularity conditions **R 1–3**, which further imply that p can be found as the unique continuously differentiable solution of the stationary forward infinitesimal Kolmogorov equations (see, for example, Reference 15, Theorems 1.9 and 2.3). This solution is called the stationary probability density function of X .

Theorem 5.2

Under conditions **PFC 1–3** and **R 1–3** the limiting probability density function p of X is given by

$$p(H, s) = cK(H) \prod_{h \notin H} [1 - F_h(s_h)] \prod_{h \in H} [1 - G_h(s_h)] \quad (5.1)$$

with c a normalizing constant and $K(H)$ given by (3.1).

Proof

First we derive the forward infinitesimal Kolmogorov equations (5.2), which express the distribution of X at some arbitrary epoch t in terms of the distribution of X at $t - \Delta$. Then, using the fact that a unique stationary probability density p for X exists which is continuously differentiable w.r.t. the residual amounts of work/repair s_h , we obtain the stationary forward infinitesimal Kolmogorov (or global balance) equations (5.6). The proof is completed by verifying that the density p given in (5.1) is indeed continuously differentiable and satisfies these global balance equations.

In order to derive the forward infinitesimal Kolmogorov equations we define for any configuration $H \in \mathcal{H}$ and for any $\Delta > 0$ the vectors $\bar{\Delta}^H$, $\underline{\Delta}^H$ and Δ^H in \mathcal{S} by

$$\bar{\Delta}_h^H = \begin{cases} \Delta\beta(h | H), & \text{if } h \notin H \\ \Delta\delta(h | H), & \text{if } h \in H \end{cases} \quad \underline{\Delta}_h^H = \begin{cases} \Delta, & \text{if } \bar{\Delta}_h^H = 0 \\ 0, & \text{if } \bar{\Delta}_h^H > 0 \end{cases}$$

and $\Delta^H = \bar{\Delta}^H + \underline{\Delta}^H$. Further we denote by e_h the h th N -dimensional unit vector. The interpretation of the vector $\bar{\Delta}^H$ is that in an infinitesimal time Δ a component $h \notin H$ performs an amount of work $\bar{\Delta}_h^H$ and a component $h \in H$ receives an amount of repair $\bar{\Delta}_h^H$. The additional vectors $\underline{\Delta}^H$ and Δ^H are introduced in order to define infinitesimal elements $(s - \Delta^H, s) \subset \mathcal{S}$ with positive volume. Consider an arbitrary state $(H, s) \in \mathcal{H} \times \mathcal{S}$ with $s_h > 0$ for all $h \in C$. As all speeds β and δ are finite we can choose Δ small enough such that $s_h - \Delta_h^H > 0$ for all $h \in C$ and thus $(H, (s - \Delta^H, s)) \subset \mathcal{H} \times \mathcal{S}$. Conditioning upon the epoch

$t - \Delta$ there are three possibilities for the event $X(t) \in (H, (s - \Delta^H, s))$ to occur:

1. The configuration does not change in the time interval $(t - \Delta, t)$, i.e.

$$X(t - \Delta) \in (H, (s - \underline{\Delta}^H, s + \bar{\Delta}^H))$$

2. The configuration changes from $H + h$ to H in $(t - \Delta, t)$ due to the repair of some component h with a residual amount of repair $\tau < \Delta\delta(h | H + h)$ at time $t - \Delta$. In this case the change in configuration occurs at time $t - \Delta + \tau/\delta(h | H + h)$. In order to account for the change of speeds in the interval $(t - \Delta, t)$ we define the vector $a(h, \tau) \in \mathcal{S}$ by

$$a_{h'}(h, \tau) = \begin{cases} \tau, & h' = h \\ s_{h'} + \tau \frac{\beta(h' | H + h) - \beta(h' | H)}{\delta(h | H + h)}, & h' \notin H + h \\ s_{h'} + \tau \frac{\delta(h' | H + h) - \delta(h' | H)}{\delta(h | H + h)}, & h' \in H \end{cases}$$

Note that, for example,

$$\tau \frac{\delta(h' | H + h) - \delta(h' | H)}{\delta(h | H + h)} + \bar{\Delta}_h^H$$

is the amount of repair received by component $h' \in H$ in the interval $(t - \Delta, t)$. Thus, given some $\tau \in (0, \Delta\delta(h | H + h))$ we must have that

$$X(t - \Delta) \in (H + h, (a(h, \tau) - \underline{\Delta}^H, a(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))$$

and the amount of work that component h can perform after its repair must be in the interval

$$\left[s_h - \tau \frac{\beta(h | H)}{\delta(h | H + h)}, s_h + \Delta_h^H - \tau \frac{\beta(h | H)}{\delta(h | H + h)} \right]$$

3. The configuration changes from $H - h$ to H in $(t - \Delta, t)$ due to the breakdown of some component h with a residual quantity of work $\tau < \Delta\beta(h | H - h)$ at time $t - \Delta$. Defining the vector $b(h, \tau) \in \mathcal{S}$ analogous to the vector $a(h, \tau)$ in the previous case, we must have that

$$X(t - \Delta) \in (H - h, (b(h, \tau) - \underline{\Delta}^H, b(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))$$

and the amount of repair that component h needs after its breakdown must be in some interval with width Δ_h^H around s_h .

All other contributions would require two or more components to change status in the interval $(t - \Delta, t)$. Owing to the regularity assumptions, these contributions will become negligible as $\Delta \downarrow 0$ and can thus be omitted. From the above observations and using the Markov

property it follows that

$$\begin{aligned}
P(X(t) \in (H, s - \Delta^H, s)) &= P(X(t - \Delta) \in (H, (s - \underline{\Delta}^H, s + \bar{\Delta}^H))) \\
&+ \sum_{\substack{h \in H: \\ H+h \in \mathcal{X}, \\ \delta(h|H+h) > 0}} \int_{\tau=0}^{\Delta\delta(h|H+h)} [f_h(s_h)\Delta_h^H + o(\Delta_h^H)] dP(X(t - \Delta) \\
&\in (H + h, (a(h, \tau) - \underline{\Delta}^H - \tau e_h, a(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))) \\
&+ \sum_{\substack{h \in H: \\ H-h \in \mathcal{X}, \\ \beta(h|h-h) > 0}} \int_{\tau=0}^{\Delta\beta(h|H-h)} [g_h(s_h)\Delta_h^H + o(\Delta_h^H)] dP(X(t - \Delta) \\
&\in (H - h, (b(h, \tau) - \underline{\Delta}^H - \tau e_h, b(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))) \quad (5.2)
\end{aligned}$$

where we have used the assumption that the random variables W_h and R_h have densities f_h respectively g_h and $o(\cdot)$ denotes any function with the property that $o(x)/x \rightarrow 0$ as $x \downarrow 0$.

Now, without loss of generality, assume that a unique stationary probability density p exists which is continuously differentiable in its residual amounts of work/repair s_h . Then, using the mean value theorem we obtain

$$\begin{aligned}
P(X(t) \in (H, (s - \Delta^H, s))) - P(X(t - \Delta) \in (H, (s - \underline{\Delta}^H, s + \bar{\Delta}^H))) \\
= \int_{\sigma \in (0, \Delta^H)} [p(H, s - \Delta^H + \sigma) - p(H, s - \underline{\Delta}^H + \sigma)] d\sigma \\
= p(H, s - \Delta^H + \sigma^1) - p(H, s - \underline{\Delta}^H + \sigma^1) \\
= - \left(\prod_{h \in C} \Delta_h^H \right) \left\{ \sum_{h \in C} \bar{\Delta}_h^H \frac{\partial}{\partial x_h} p(H, x) \Big|_{s - \underline{\Delta}^H + \sigma^1} + o(\|\Delta^H\|) \right\} \quad (5.3)
\end{aligned}$$

for some vector $\sigma^1 \in (0, \Delta^H)$. Note that $s - \Delta^H + \sigma^1 \rightarrow s$ along some trajectory in the interior of \mathcal{S} as $\Delta \downarrow 0$. Again using the mean value theorem, for each separate integral in the second term of the r.h.s. of (5.2) we obtain (apart from the factor $f_h(s_h)\Delta_h^H + o(\Delta_h^H)$)

$$\begin{aligned}
\int_{\tau=0}^{\Delta\delta(h|H+h)} dP(X(t - \Delta) \in (H + h, (a(h, \tau) - \underline{\Delta}^H - \tau e_h, a(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))) \\
= \int_{\tau=0}^{\Delta\delta(h|H+h)} \int_{\sigma \in (0, \Delta^H - \Delta_h^H e_h)} p(H + h, a(h, \tau) - \underline{\Delta}^H + \sigma) d\sigma d\tau \\
= \left(\prod_{\{h' \in C: h' \neq h\}} \Delta_{h'}^H \right) p(H + h, a(h, \tau^2) - \underline{\Delta}^H + \sigma^2) \Delta\delta(h|H+h) \quad (5.4)
\end{aligned}$$

for some value $\tau^2 \in (0, \Delta\delta(h|H+h))$ and vector $\sigma^2 \in (0, \Delta^H - \Delta_h^H e_h)$. In a similar manner we obtain for each integral in the third term of the r.h.s. of (5.2) (apart from the factor $g_h(s_h)\Delta_h^H + o(\Delta_h^H)$)

$$\begin{aligned}
\int_{\tau=0}^{\Delta\beta(h|H-h)} dP(X(t - \Delta) \in (H - h, (b(h, \tau) - \underline{\Delta}^H - \tau e_h, b(h, \tau) + \bar{\Delta}^H - \bar{\Delta}_h^H e_h))) \\
= \left(\prod_{\{h' \in C: h' \neq h\}} \Delta_{h'}^H \right) p(H - h, b(h, \tau^3) - \underline{\Delta}^H + \sigma^3) \Delta\beta(h|H-h) \quad (5.5)
\end{aligned}$$

for some value $\tau^3 \in (0, \Delta\beta(h|H-h))$ and vector $\sigma^3 \in (0, \Delta^H - \Delta_h^H e_h)$. Note that both $a(h, \tau^2) - \underline{\Delta}^H + \sigma^2 \rightarrow s - s_h e_h$ and $b(h, \tau^3) - \underline{\Delta}^H + \sigma^3 \rightarrow s - s_h e_h$ along some trajectory

in the interior of \mathcal{S} as $\Delta \downarrow 0$. Finally, eliminating the common factor $\prod_{h' \in C} \Delta_{h'}^H$ in (5.3)–(5.5), dividing by Δ and letting $\Delta \downarrow 0$ we obtain the stationary forward infinitesimal Kolmogorov equations

$$\begin{aligned} & \left\{ \sum_{\substack{h \notin H: \\ H+h \in \mathcal{H}, \\ \delta(h|H+h) > 0}} \right\} \lim_{x \rightarrow s - s_h e_h} p(H+h, x) \delta(h|H+h) f_h(s_h) - \left\{ \sum_{\substack{h \notin H: \\ \beta(h|H) > 0}} \right\} \frac{\partial}{\partial x_h} p(H, x) \Big|_s \beta(h|H) \\ & + \left\{ \sum_{\substack{h \in H: \\ H-h \in \mathcal{H}, \\ \beta(h|H-h) > 0}} \right\} \lim_{x \rightarrow s - s_h e_h} p(H-h, x) \beta(h|H-h) g_h(s_h) - \left\{ \sum_{\substack{h \in H: \\ \delta(h|H) > 0}} \right\} \frac{\partial}{\partial x_h} p(H, x) \Big|_s \delta(h|H) \\ & = 0 \end{aligned} \tag{5.6}$$

All that remains is to verify that the probability density p as given in (5.1) above satisfies these equations. Consider some component $h \notin H$ with $\beta(h|H) > 0$. First note that the regularity condition **R 3** and **PFC 2** imply that $H+h \in \mathcal{H}$ and $\delta(h|H+h) > 0$. Differentiating (5.1) yields

$$\frac{\partial}{\partial x_h} p(H, x) \Big|_s = -f_h(s_h) cK(H) \prod_{h \notin H+h} [1 - F_h(s_h)] \prod_{h \in H} [1 - G_h(s_h)] \tag{5.7}$$

and using $\lim_{x_h \downarrow 0} [1 - F_h(x_h)] = 1$ we obtain

$$\lim_{x \rightarrow s - s_h e_h} p(H+h, x) = cK(H+h) \prod_{h \notin H+h} [1 - F_h(s_h)] \prod_{h \in H} [1 - G_h(s_h)] \tag{5.8}$$

With condition **PFC 3'** it follows that all contributions due to components $h \notin H$ in (5.6) are equal to zero. In a similar manner it can be shown that all contributions due to components $h \in H$ are equal to zero also. □

Using Theorem 5.2 the steady state probability $\pi(H)$ that the system is in configuration H can be readily obtained by integrating the limiting probability density function p over all possible residual amounts s_h of work (if $h \notin H$) or repair (if $h \in H$).

Corollary 5.3

Under conditions **PFC 1–3** and **R 1–3** we have for all $H \in \mathcal{H}$ that

$$\pi(H) = \bar{c}K(H) \prod_{h \in H} \frac{E[R_h]}{E[W_h]} \tag{5.9}$$

with \bar{c} a normalizing constant and $K(H)$ given by (3.1).

Proof

The change of order of integration and taking products below is justified by the fact that all integration variables occur in exactly one separate factor in the product form (5.1). Thus,

$$\begin{aligned} \pi(H) &= \int_{\mathcal{S}} p(H, s) ds = cK(H) \prod_{h \notin H} \int_0^\infty [1 - F_h(s_h)] ds_h \prod_{h \in H} \int_0^\infty [1 - G_h(s_h)] ds_h \\ &= cK(H) \prod_{h \notin H} E[W_h] \prod_{h \in H} E[R_h] = \bar{c}K(H) \prod_{h \in H} \frac{E[R_h]}{E[W_h]} \end{aligned} \tag{5.10}$$

where $\bar{c} = c \prod_{h=1}^N E[W_h]$. □

6. EXAMPLES

In this section we give a number of examples that illustrate the kind of systems that can be modelled, i.e. for which the speed functions satisfy the product form conditions.

6.1. Simple series-parallel system

Consider a system consisting of one critical component, a , and a set $B = \{b_1, \dots, b_M\}$ of M secondary components (see Figure 2). The system is operational if the critical component and at least $M - k$ of the secondary components are up. We assume that none of the components can be used if the system is not operational, so that the set of possible configurations is given by

$$\mathcal{H} = \{H \subset C : |H| \leq k + 1\}$$

Further we assume that there is a single repair unit giving pre-emptive priority to the critical component and that the secondary components are repaired in a processor sharing manner. Defining $n_B(H) = |H \cap B|$ as the number of secondary components that are down in configuration H , the breakdown and repair speeds for the secondary components $b \in B$ are given by

$$\beta(b|H) = \begin{cases} 1, & \text{if } a \notin H \text{ and } |H| \leq k - 1 \\ 0, & \text{else} \end{cases} \quad \delta(b|H) = \begin{cases} \frac{1}{n_B(H)}, & \text{if } a \notin H \\ 0, & \text{else} \end{cases}$$

and for the critical component a we have

$$\beta(a|H) = \begin{cases} 1, & \text{if } n_B(H) \leq k - 1 \\ 0, & \text{else} \end{cases} \quad \delta(a|H) = 1$$

It can easily be seen that **PFC 1** and **PFC 2** are satisfied. In order to check **PFC 3'**, the equivalent of **PFC 3**, observe that if $a \in H$ all repair paths $\gamma \in \Gamma_\delta(H)$ have $\gamma(1) = a$, i.e. the critical component is repaired first. Thus, the only conditions for the function K are

$$K(H) = \begin{cases} K(H - a) \frac{\beta(a|H - a)}{\delta(a|H)} = K(H - a), & \text{if } a \in H \\ K(H - b) \frac{\beta(b|H - b)}{\delta(b|H)} = K(H - b)n_B(H) \text{ for all } b \in H \cap B, & \text{if } a \notin H \end{cases}$$

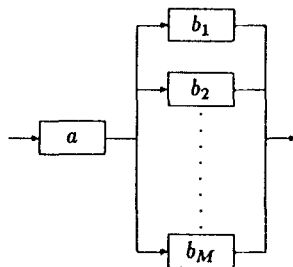


Figure 2. Series-parallel system

As a consequence K is given by $K(H) = n_B(H)!$ and from Corollary 5.3 we conclude that

$$\pi(H) = cn_B(H)! \prod_{h \in H} \frac{E[R_h]}{E[W_h]} \tag{6.1}$$

with c a normalizing constant.

Remark 6.1

From Section 4.2 it can be concluded that instead of a processor sharing discipline, also an FIFO queue for the repair of the secondary components ($Q = B$) leads to a product form steady state solution provided that for all $h \in B$ the required quantity of repair R_h is exponentially distributed with mean μ^{-1} . The corresponding function δ for $b \in B$ in (4.6) is given by

$$\delta(b | H) = \begin{cases} 1, & \text{if } a \notin H \\ 0, & \text{else} \end{cases}$$

Otherwise the functions β and δ are as given above. It can easily be verified that conditions **Q 1–3** and **PFC 1–3** are satisfied, now with $K(H) = 1$, so that from Theorem 4.2 we conclude that the steady state probabilities $\bar{\pi}(H, \xi)$ are given by

$$\bar{\pi}(H, \xi) = c \left(\prod_{h \in H \cap B^c} \frac{E[R_h]}{E[W_h]} \right) \left(\prod_{h \in H \cap B} \frac{\mu^{-1}}{E[W_h]} \right)$$

with c a normalizing constant. Noting that $\pi(H, \xi)$ does not depend on ξ and that there are $n_B(H)!$ possible orderings ξ we see that the steady state probability that the set of components H is down is still given by (6.1), of course with $E[R_h] = \mu^{-1}$ for $h \in B$.

6.2. Fault tolerant database system

Consider a database system consisting of a front-end, two processing subsystems and a database. Each subsystem consists of a switch, a memory and two processors and is operational if the switch, the memory and at least one of its processors are up. The system as a whole is operational if the front-end and the database are up and at least one of the processing subsystems is operational. The system can be interpreted as a nested series-parallel system (see Figure 3).

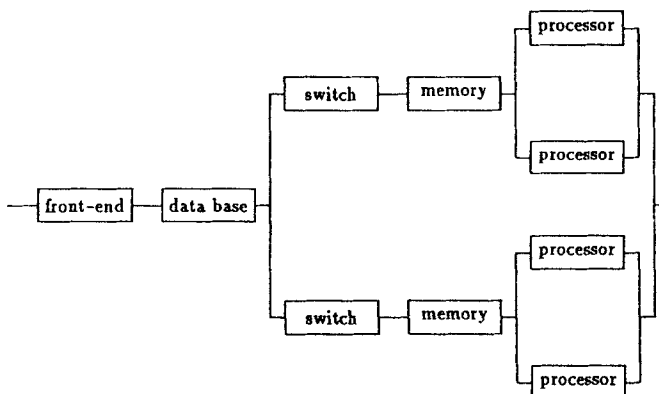


Figure 3. Nested series-parallel system

We assume that the components in a subsystem cannot be used if that subsystem is down and if the whole system is down none of the components can be used. In order to define the speed functions and the set of all possible configurations we define for each component $h \in C$ its direct dependability set $D(h)$ as the set of components, other than h , that can only be used if component h is up. For example, if h denotes a processor $D(h) = \emptyset$ and if h denotes a switch, $D(h)$ contains all the component of the subsystem, except for the switch itself. Further, from Figure 3 it is apparent that the system has four different path sets $P_i \subset C$, $i = 1, \dots, 4$, where a path set is defined as a minimal set of components that must be up for the system as a whole to be operational. Let us label the four processors by p_1, \dots, p_4 such that $p_i \in P_i$, $i = 1, \dots, 4$. The set of all possible configurations \mathcal{H} is given by

$$\mathcal{H} = \left\{ H \subset C: \min_i \{|H \cap D(h) \cap P_i|\} = 0 \text{ for all } h \in H \right\}$$

i.e. each component was contained in at least one path set with all components up just before it broke down. For the breakdown speeds, we have for all $H \in \mathcal{H}$ and all $h \notin H$

$$\beta(h | H) = 0 \Leftrightarrow \exists h' \in H: h \in D(h') \quad (6.2)$$

As **PFC 2** requires that for all $H \in \mathcal{H}$ and all $h \in H$

$$\delta(h | H) = 0 \Leftrightarrow \exists h' \in H: h \in D(h') \quad (6.3)$$

we must assume a hierarchic repair strategy, which we model as follows. Each subsystem has its own repair unit, giving pre-emptive priority to the memory and the switch. (Note that, within a subsystem, the memory and the switch cannot be down at the same time.) If the database goes down then *both* repair units are assigned to repair the database immediately. The same holds for the front end. Thus, the database and the front end have pre-emptive priority at system level. The repair units resume the work they were doing as soon as they finish the repair of a component with higher priority, i.e. the priority disciplines are work conserving.

Having modelled the hierarchy, i.e. when speeds equal zero, we have the freedom to assign the positive speeds in an arbitrary manner. Suppose that the repair speed for a processor is equal to 1 if it is the only processor that is down in its subsystem but that, if the other one goes down too, this speed drops to q for both processors. Further, if in a subsystem both processors are up then they both work at speed 1 and if only one of them is up then it works at speed p . The speeds of all other components take the values 0 or 1 only.

Again, it is easy to see that **PFC 1** is satisfied. That **PFC 2** is satisfied follows from (6.2) and (6.3). In order to check **PFC 3**, consider a configuration $H \neq \emptyset$ and an arbitrary repair path $\gamma \in \Gamma_\delta(H)$. The only factors

$$\frac{\beta(\gamma(k) | H_{k-1}^\gamma)}{\delta(\gamma(k) | H_k^\gamma)}$$

that matter in (3.1) are those for which $\gamma(k)$ denotes a processor, all other factors β/δ being equal to 1. From this observation it follows that the function

$$K(H) = \begin{cases} 1, & \text{if in both subsystems at most one processor is down} \\ p/q, & \text{if in one subsystem both processors are down} \\ (p/q)^2, & \text{if all four processors are down} \end{cases}$$

satisfies **PFC 3**.

As in Remark 6.1 it can be verified that the steady state distribution π is still of product form if the processors in each subsystem join a FIFO queue for repair (one separate queue for each

subsystem), provided that they require exponentially distributed amounts of repair with means μ_a^{-1} and μ_b^{-1} for subsystem a and b, respectively. Note that the case where all four processors join the same queue would violate condition Q 3.

6.3. Multi-stage interconnection network

Consider a multi-stage interconnection network (MIN) connecting K inputs to K outputs. We assume that K is a power of two. When using 2×2 internal switches to build up the MIN, there are $M = {}^2\log K$ stages required. The number of switches per stage is $N = K/2$. In Figure 4 an 8×8 MIN is depicted.

Assume that under normal operation all switches that are up work at speed $\beta = 1$. Switches that are down are repaired in a processor sharing manner. However, there is a critical number B_i , $1 \leq B_i \leq N$, associated with each stage, $i = 1, \dots, M$. As soon as B_i switches of stage i are down the network abandons normal operation. Stage i is now critical and all switches that are up stop working. The repair of all switches that are not in the critical stage is also stopped in favour of the B_i switches that are down in the critical stage i . These are repaired in a processor sharing manner. As soon as one of the switches in the critical stage is repaired the system resumes normal operation again.

Let C_i denote the set of switches in stage i and let $n_i(H) = |H \cap C_i|$ denote the number of switches that are down in stage i if the system is in configuration H , $i = 1, \dots, M$.

The set \mathcal{H} of all possible configurations is given by

$$\mathcal{H} = \{H \subset C: (\exists j \in \{1, \dots, M\})(n_j(H) \leq B_j \text{ and } n_i(H) < B_i, i \neq j)\}$$

For all states H and switches $h \in C$ the speeds β and δ are given by

$$\beta(h | H) = \begin{cases} 1, & \text{if } n_i(H) < B_i, i = 1, \dots, M \\ 0, & \text{else} \end{cases}$$

and

$$\delta(h | H) = \begin{cases} 1/|H|, & \text{if } n_i(H) < B_i, i = 1, \dots, M \\ 1/B_j, & \text{if } n_j(H) = B_j \text{ for some } j \in \{1, \dots, M\} \text{ and } h \in S_j \\ 0, & \text{if } n_j(H) = B_j \text{ for some } j \in \{1, \dots, M\} \text{ and } h \notin S_j \end{cases}$$

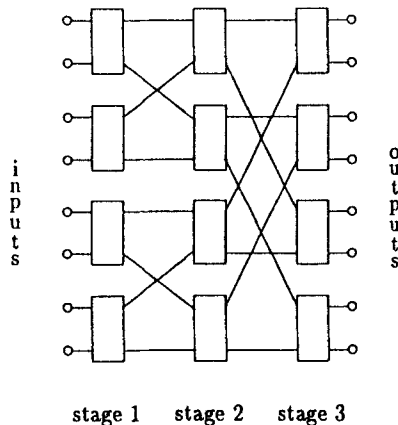


Figure 4. Multi-stage interconnection network

From this, it is not difficult to see that **PFC 1** and **PFC 2** are satisfied. Finally, in the same manner as was done in Example 6.1, it can be verified that the function $K(H)$ given by

$$K(H) = \begin{cases} |H|!, & \text{if } n_i(H) < B_i, i = 1, \dots, M \\ (|H| - 1)!B_j, & \text{if } n_j(H) = B_j \text{ for some } j \in \{1, \dots, M\} \end{cases}$$

satisfies **PFC 3'**.

7. CONCLUSION

In this paper we have characterized a class of availability models that exhibit a product form steady state solution. The conditions that have to be fulfilled for a model to fall in this class (the product form conditions) have been stated and examples are given that show how these conditions can be verified. Further research is conducted on using the product form results to obtain bounds for non-product-form models. This can be done by making proper adjustments in the speed functions β and δ .

ACKNOWLEDGEMENT

The authors are very grateful for the helpful comments of the referees.

REFERENCES

1. A. L. Reibman, R. M. Smith and K. S. Trivedi, 'Markov and Markov reward model transient analysis: an overview of numerical approaches', *Europ. J. Operations Research*, **40**, 257–267 (1989).
2. S. M. Ross, 'On the calculation of asymptotic system reliability characteristics', in *Reliability and Fault Tree Analysis*, SIAM, Philadelphia, 1975.
3. R. Addie and R. Taylor, 'An Algorithm for calculating the availability and mean time to repair for communication through a network', in *ITC Specialist Seminar*, Adelaide, Paper No. 10.5, 1989.
4. A. Agrawal and R. E. Barlow, 'A survey of network reliability and domination theory', *Operations Research*, **32**, 478–492 (1984).
5. T. B. Brecht and C. J. Colbourn, 'Lower bounds on two-terminal network reliability', *Discrete Applied Mathematics*, **21**, 185–198 (1988).
6. V. O. K. Li and J. A. Silvester, 'Performance analysis of networks with unreliable components', *IEEE Trans. Commun.*, **32**(10), 1105–1110 (1984).
7. A. Goyal, S. S. Lavenberg and K. S. Trivedi, 'Probabilistic modeling of computer systems availability', *Ann. Operations Research*, **8**, 285–306 (1987).
8. R. R. Muntz, E. de Souza e Silva and A. Goyal, 'Bounding availability of repairable computer systems', *IEEE Trans. Comput.*, **38**(12), 1714–1723 (1989).
9. N. M. van Dijk and J. P. Veltkamp, 'Product forms for interference systems', *Probability Eng. Inf. Sciences*, 355–376 (1988).
10. F. P. Kelly, *Reversibility and Stochastic Networks*, Wiley, 1979.
11. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, 1975.
12. A. Hordijk and N. M. van Dijk, 'Adjoint processes, job local balance and insensitivity for stochastic networks', *Bull. 44th Session Int. Stat. Inst.*, Vol. 50, 1983, pp. 776–788.
13. R. Schassberger, 'Insensitivity of steady-state distributions of generalised semi-Markov processes with speeds', *Adv. Appl. Probability*, **10**, 836–851 (1978).
14. P. Whittle, *Systems in Stochastic Equilibrium*, Wiley, 1986.
15. E. B. Dynkin, *Markov Processes I*, Springer-Verlag, Berlin, 1965.
16. F. Baskett, K. M. Chandy, R. R. Muntz and F. G. Palacios, 'Open, closed and mixed networks of queues with different classes of customers', *Journal of the ACM*, **22**(2), 248–260 (1975).