The machine repair model

Sander Dommers (0527269)
Bachelor Project

Supervisor: Prof.dr.ir. O.J. Boxma

March 4, 2008
Version 1.0

Contents

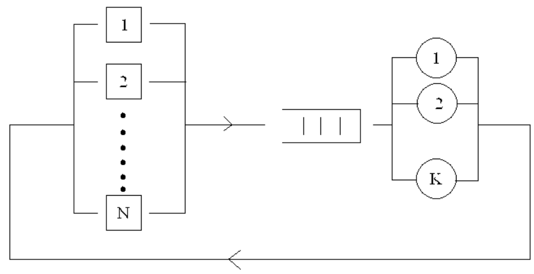

Figure 1: The machine repair model

1 Introduction

Many machines are unreliable. This means that they have the tendency to break down. These breakdowns may have severe impact on the performance of the machine and even on the performance of other machines because there might be too little repairmen. This report will investigate some models that will help analyzing this effect.

In Section 2 we will discuss some models where a finite number of machines can break down and have to be repaired by a small number of repairmen. We will discuss both the equilibrium distribution of the number of machines that is down, as well as the downtime distribution of an arbitrary machine that breaks down. Also the impact of different repair strategies is discussed.

In Section 3 we will study a model by Wartenhorst, which is the same as the model in Section 2, but now the machines themselves also have to serve jobs and all machines have their own queue. We use the behavior that is derived for the downtime of the machines in Section 2 to study the equilibrium distribution of the number of jobs in a certain queue.

Next, in Section 4 we study a model by Mitrani et al. where we will assume that the machines break down and are repaired independently of each other. The machines do all have their own queue, but this time there is only one stream of arriving jobs. These jobs have to be routed to the different machines, where the routing may depend on what machines are down. Two different models are studied. In Section 4.1 the jobs that are in a queue of a machine that breaks down are lost and in Section 4.2 the jobs remain in place when a machine breaks down.

Finally, in Section 5 we will study one unreliable machine which serves its jobs according to the Processor Sharing discipline. This machine can only serve a finite, N say, number of jobs. All jobs that arrive when there are already N jobs in the system will be lost. We will analyze both the equilibrium distribution of the number of jobs in the system as well as the sojourn time of an arbitrary job. A comparison is made with the First Come, First Serve discipline and it is discussed how the system can be analyzed when the downtime has a general distribution.

2 The machine repair model

2.1 Model description

Consider N machines that are subject to breakdowns. Let there be K repairmen to repair these machines. In a large part of the report we'll take $K = 1$. If $K < N$, then machines may have to wait before the repair starts. This model is shown in Figure 1.

Assume the lifetimes of the machines are i.i.d. (independent and identically distributed) with mean $1/\sigma$. The successive repair times are assumed to be i.i.d. with mean $1/\nu$. We consider several repair strategies: Repair in a First Come, First Serve (FCFS) order, in Random Order of Service (ROS) or according to the Processor Sharing (PS) discipline.

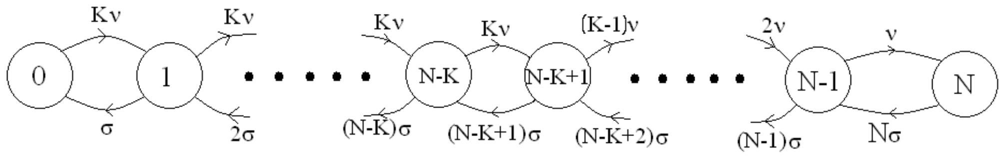

Figure 2: A view of the continuous-time Markov chain of the number of working machines

2.2 Queue length analysis

2.2.1 Memoryless lifetimes and repair times

Let the lifetimes of the machines be $Exp(\sigma)$ distributed and the repair times $Exp(\nu)$. Let all the lifetimes and repair times be independent of each other. Due to the exponential repair times, the order in which the machines are served does not matter. Let $X(t)$ be the number of working machines. Then $\{X(t)|t \geq 0\}$ is a continuous-time Markov Chain. This chain is depicted in Figure 2.

The limiting probabilities $p_n = \lim_{t\to\infty} P[X(t) = n]$ exist because all states of the Markov chain communicate and are positive recurrent, due to the finite number of states. The limiting probabilities can be computed with the following balance equations:

$$\begin{aligned} K\nu p_n &= (n+1)\sigma p_{n+1}, \qquad 0 \leq n \leq N-K, & (1)\\ (N-n)\nu p_n &= (n+1)\sigma p_{n+1}, \qquad N-K \leq n < N. & (2) \end{aligned}$$

All $p_n, 0 \leq n \leq N$, can be expressed in p_0 as follows:

$$\begin{aligned} p_n &= \frac{1}{n!}K^n(\frac{\nu}{\sigma})^n p_0, \qquad 0 \leq n \leq N-K+1, & (3)\\ p_n &= \frac{1}{n!}K^{N-K}\frac{K!}{(N-n)!}(\frac{\nu}{\sigma})^n p_0, \qquad N-K+1 \leq n \leq N. & (4) \end{aligned}$$

Because

$$\sum_{n=0}^{N} p_n = 1, \tag{5}$$

it can be concluded that

$$p_0 = \left(\sum_{n=0}^{N-K+1}\frac{1}{n!}K^n(\frac{\nu}{\sigma})^n + \sum_{n=N-K+1}^{N}\frac{1}{n!}K^{N-K}\frac{K!}{(N-n)!}(\frac{\nu}{\sigma})^n\right)^{-1}. \tag{6}$$

2.2.2 Two-phase memoryless lifetimes and memoryless repair times

Let the lifetimes of the machines be the sum of two independent phases which are exponential distributed with mean $1/\sigma_1$ and $1/\sigma_2$ and let the repair times be $Exp(\nu)$ distributed. Let these repair times be independent of each other and independent of the lifetimes and let there be one repairman, i.e. $K = 1$. Consider the two-dimensional Markov chain $\{(X(t), Y(t))|t \geq 0\}$ where $X(t)$ is the number of machines in their first phase and $Y(t)$ the number of machines in their second phase. This means that at time t, $N - X(t) - Y(t)$ machines are broken.

For $i > 0, j > 0$ and $i + j < N$, the limiting probabilities $p_{i,j} = \lim_{t\to\infty} P[(X(t), Y(t)) = (i, j)]$ satisfy the following balance equations:

$$(i\sigma_1 + j\sigma_2 + \nu)p_{i,j} = \nu p_{i-1,j} + (i+1)\sigma_1 p_{i+1,j-1} + (j+1)\sigma_2 p_{i,j+1}. \tag{7}$$

To solve these equations we try

$$p_{i,j} = C\frac{(\frac{\nu}{\sigma_1})^i}{i!}\frac{(\frac{\nu}{\sigma_2})^j}{j!}. \tag{8}$$

This guess satisfies equation (7), because

$$
\begin{aligned}
\frac{p_{i-1,j}}{p_{i,j}} &= \frac{i\sigma_1}{\nu}, && (9)\\
\frac{p_{i+1,j-1}}{p_{i,j}} &= \frac{j\sigma_2}{(i+1)\sigma_1}, && (10)\\
\frac{p_{i,j+1}}{p_{i,j}} &= \frac{\nu}{(j+1)\sigma_2}. && (11)
\end{aligned}
$$

If $i = 0$, then $p_{i-1,j} = 0$, if $j = 0$, then $p_{i+1,j-1} = 0$ and if $i + j = N$ then $p_{i,j+1} = 0$ and the term $\nu p_{i,j}$ also vanishes from equation (7). So it can be concluded that this guess also satisfies the balance equation in all other cases.

C can now be computed as

$$C = \left(\sum_{i=0}^{N} \sum_{j=0}^{N-i} \frac{(\frac{\nu}{\sigma_1})^i}{i!} \frac{(\frac{\nu}{\sigma_2})^j}{j!} \right)^{-1}. \tag{12}$$

If we want to know the probability that in the long run n machines are working, we can say

$$
\begin{aligned}
\lim_{t\to\infty} P[n \text{ machines are working}] &= \sum_{i=0}^{n} p_{i,n-i}\\
&= C \sum_{i=0}^{n} \frac{(\frac{\nu}{\sigma_1})^i}{i!} \frac{(\frac{\nu}{\sigma_2})^{n-i}}{(n-i)!}\\
&= \frac{C}{n!} \sum_{i=0}^{n} (\frac{\nu}{\sigma_1})^i (\frac{\nu}{\sigma_2})^{n-i} \binom{n}{i}\\
&= \frac{C}{n!} (\frac{\nu}{\sigma_1} + \frac{\nu}{\sigma_2})^n. && (13)
\end{aligned}
$$

Thus, if $1/\sigma_1 + 1/\sigma_2 = 1/\sigma$ this gives the same result as in Section 2.2.1. This means that only the first moment of the lifetimes of the machines is important.

2.2.3 $M/G/N$ loss model

Our machine repair model with only one repairman, $K = 1$, is the same as an Erlang loss model with N machines, where the arriving rate of jobs is ν, the distribution of processing times of jobs is the same as the lifetime distribution of the machines with mean $1/\sigma$ and there are N machines (see Figure 3). This means that working machines arrive with rate ν when not all N machines are working and with rate 0 if all machines are working (additional arriving working machines are blocked). Let $\rho = \frac{\sigma}{\nu}$. In [2] Cohen studies the Erlang loss model and shows that for the $M/G/N$ loss model Erlang's formula

$$\lim_{t\to\infty} P[X(t) = n] = \frac{\rho^n}{n!} \frac{1}{1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^N}{N!}}, \qquad n = 0, 1, \ldots, N, \tag{14}$$

holds, where $X(t)$ is the number of jobs in the system, N is the number of machines and ρ is the amount of work offered per unit of time. So the steady state distribution of the number of working machines is *insensitive* to the lifetime distribution, apart from its mean.

The proof that the stationary distribution of the Erlang loss model is indeed given by Erlang's formula consists of three stages. First Cohen shows that Erlang's formula holds for an $M/M/\infty$ queue, which is a well known result.

Second he proves that this formula also holds for an $M/E/\infty$ queue, where E is an Erlang distribution, and thereafter also proves that this formula holds when E is any finite mixture of

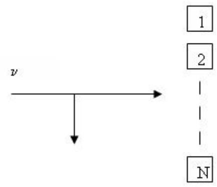

Figure 3: The M/G/N loss model

Erlang distributions. He concludes with Schassberger [9] that this result can be generalized to all $M/G/\infty$ models, where G is a general distribution.

Third we have to prove that Erlang's formula also holds for the $M/G/N$ loss model. To do this, Cohen starts to look at the imbedded Markov process $\{(\widehat{X}^{(\infty)}(m), \underline{\zeta}(m)) | m = 1, 2, ...\}$, where $\widehat{X}^{(\infty)}(m)$ is the number of jobs being processed in the $M/G/\infty$ model immediately after the m-th departure and $\underline{\zeta}(m)$ are the residual service times of the $\widehat{X}^{(\infty)}(m)$ jobs. For this system he formulates the one-step transition probabilities and, using these, the system of forward relations. It is well known that the solution of this system satisfies (14). He then does the same for the process $\{(\widehat{X}^{(N)}(m), \underline{\zeta}(m)) | m = 1, 2, ...\}$, where $\widehat{X}^{(N)}(m)$ also is the number of jobs in the system immediately after the m-th departure, but now in the $M/G/N$ loss model, where a job that finds all N machines busy is also counted as a departure. Cohen shows that the forward relations for these two systems are the same up to a constant that only depends on N. This proves that $\lim_{m\to\infty} P[\widehat{X}^{(N)}(m) = n]$ equals the righthand side of (14).

Because the busy cycles of the $M/G/N$ loss model are finite, the number of times an arriving job sees n jobs being processed must be the same as the number of times a departing job leaves n jobs behind when leaving. Let $\mathbb{I}_{\{A\}}$ be the indicator function that the event A is true. Then

$$E[\sum_{i=1}^{i^{(N)}} \mathbb{I}_{\{X^{(N)}(t_{i-})=n\}}] = E[\sum_{m=1}^{i^{(N)}} \mathbb{I}_{\{X^{(N)}(r_{m+})=n\}}], \qquad n = 0, 1, \ldots, N, \tag{15}$$

with $i^{(N)}$ the number of jobs in a busy cycle, t_{i-} the time just before the i-th job arrives and r_{m+} the time just after the m-th job leaves. Now we can conclude with (15) and the PASTA property [13] that

$$\begin{aligned} \lim_{m\to\infty} P[\widehat{X}^{(N)}(m) = n] &= \frac{1}{E[i^{(N)}]} E[\sum_{m=1}^{i^{(N)}} \mathbb{I}_{\{X^{(N)}(r_{m+})=n\}}] \\ &= \frac{1}{E[i^{(N)}]} E[\sum_{i=1}^{i^{(N)}} \mathbb{I}_{\{X^{(N)}(t_{i-})=n\}}] \\ &= \lim_{i\to\infty} P[X^{(N)}(t_{i-}) = n] \\ &= \lim_{t\to\infty} P[X^{(N)}(t) = n]. \end{aligned} \tag{16}$$

This shows that the $M/G/N$ loss model satisfies Erlang's formula.

2.3 Downtime analysis

In the following sections we study the length of an arbitrary downtime of a machine D, when the machines are repaired according to a given discipline. We assume the lifetimes of the machines are i.i.d. and are $Exp(\sigma)$ distributed. The repair times are independent of the lifetimes and they themselves are i.i.d. and $Exp(\nu)$ distributed.

2.3.1 FCFS

Assume that the machines are repaired in FCFS order. The distribution of D is obtained by conditioning on the number of broken machines found by a machine that just broke down (the tagged machine).

$$r_j \equiv P(\text{tagged machine finds } j \text{ machines at the repair facility}), \qquad 0 \le j < N. \tag{17}$$

The N machines are jobs moving around in a closed queueing network. The Arrival Theorem of Lavenberg & Reiser [4] states that the stationary state probabilities at instants at which jobs move from one service station to another are equal to the stationary state probabilities at a random point in time for the network with one less job. Thus, r_j can be computed as p_{N-j} in Section 2.2 with N replaced by $N-1$.

If the tagged machine finds j machines that have to be repaired, with $0 \le j < K$, a repairman is available so the tagged machine leaves the repair facility in an $Exp(\nu)$ amount of time. If $K \le j < N$ the tagged machine has to wait until $j-K+1$ machines are repaired. The time it takes until one machine is repaired is the minimum of K repair times that are $Exp(\nu)$ distributed, so this time is $Exp(K\nu)$ distributed. The time it takes until $j-K+1$ machines are repaired is thus the sum of $j-K+1$ stochastic variables that are $Exp(K\nu)$ distributed, so the machine has to wait an $Erlang(j-K+1, K\nu)$ amount of time. Then the machine itself has to be repaired, so there is an additional downtime that is $Exp(\nu)$ distributed. So the Laplace-Stieltjes transform is given by

$$\begin{aligned} E[e^{-sD}] &= \sum_{j=0}^{N-1} r_j E[e^{-sD} | \text{tagged machine finds } j \text{ machines at the repair facility}] \\ &= \sum_{j=0}^{K-1} r_j \frac{\nu}{\nu+s} + \sum_{j=K}^{N-1} r_j (\frac{K\nu}{K\nu+s})^{j-K+1} \frac{\nu}{\nu+s}. \end{aligned} \tag{18}$$

The computation of the moments of D now is simple. The first two moments, for example, are given by

$$E[D] = \sum_{j=0}^{K-1} r_j \frac{1}{\nu} + \sum_{j=K}^{N-1} r_j (\frac{j-K+1}{K\nu} + \frac{1}{\nu}), \tag{19}$$

$$E[D^2] = \sum_{j=0}^{K-1} r_j \frac{2}{\nu^2} + \sum_{j=K}^{N-1} r_j (\frac{j-K+1+(j-K+1)^2}{(K\nu)^2} + \frac{j-K+1}{K\nu}\frac{1}{\nu} + \frac{2}{\nu^2}). \tag{20}$$

2.3.2 ROS

Make the same assumptions as above, but now assume the machines are repaired in Random Order of Service. The distribution of D again is obtained by conditioning on the number of machines found by a machine that just broke down (the tagged machine). r_j is defined as above and can be computed as described there.

If the tagged machine finds j machines that have to be repaired, with $0 \le j < K$, a repairman is available so the tagged machine leaves the repair facility in an $Exp(\nu)$ amount of time. If $K \le j < N$, then after an $Exp(K\nu + (N-j-1)\sigma)$ amount of time an event happens: with probability $\frac{K\nu}{K\nu+(N-j-1)\sigma}$ a machine has been repaired and with probability $\frac{(N-j-1)\sigma}{K\nu+(N-j-1)\sigma}$ another machine breaks down. In the first case $j-K+1$ machines are waiting to be repaired. With probability $\frac{1}{j-K+1}$ the tagged machine goes into service and thus remains down for an $Exp(\nu)$ amount of time. Otherwise an other machine goes into service and it is like the tagged machine arrives and finds one less machine in the repair facility. In the second case it is like the tagged machine arrives and finds one more machine in the repair facility.

If we define G_j as the amount of time a machine stays in the repair facility given he finds j machines when he just breaks down, and $G_j^*(s)$ as the Laplace-Stieltjes transform of G_j, then we find the following recurrence relation of $G_j^*(s)$

$$G_j^*(s) = \begin{cases} \frac{\nu}{\nu+s}, & 0 \le j < K, \\ \frac{K\nu+(N-j-1)\sigma}{K\nu+(N-j-1)\sigma+s}\left(\frac{K\nu}{K\nu+(N-j-1)\sigma}\left(\frac{1}{j-K+1}\frac{\nu}{\nu+s}+\frac{j-K}{j-K+1}G_{j-1}^*(s)\right)\right. \\ \qquad \left.+\frac{(N-j-1)\sigma}{K\nu+(N-j-1)\sigma}G_{j+1}^*(s)\right), & K \le j < N. \end{cases} \tag{21}$$

These equations can be easily solved, but this doesn't give a nice explicit formula. The moments can be obtained by differentiating and then can be fairly easy numerically evaluated.

2.3.3 PS

If machines are repaired according to the Processor Sharing discipline, almost the same arguments as with ROS hold. The first event after a machine breaks down that finds j machines happens after an $Exp(\min\{j+1,K\}\nu+(N-j-1)\sigma)$ amount of time. If the first event is a departure of a machine, with probability $\frac{1}{j+1}$ this was the tagged machine. Otherwise it is just like the tagged machine arrives and finds $j-1$ machines. If the first event after arrival of the tagged machine is an arrival of a machine it is just like the tagged machine arrives and finds $j+1$ machines at the repair facility. A recurrence relation for $G_j^*(s)$ now is given by

$$G_j^*(s) = \begin{cases} \frac{(j+1)\nu+(N-j-1)\sigma}{(j+1)\nu+(N-j-1)\sigma+s}\left(\frac{(j+1)\nu}{(j+1)\nu+(N-j-1)\sigma}\left(\frac{1}{j+1}1+\frac{j}{j+1}G_{j-1}^*(s)\right)\right. \\ \qquad \left.+\frac{(N-j-1)\sigma}{(j+1)\nu+(N-j-1)\sigma}G_{j+1}^*(s)\right), & 0 \le j < K, \\ \frac{K\nu+(N-j-1)\sigma}{K\nu+(N-j-1)\sigma+s}\left(\frac{K\nu}{K\nu+(N-j-1)\sigma}\left(\frac{1}{j+1}1+\frac{j}{j+1}G_{j-1}^*(s)\right)\right. \\ \qquad \left.+\frac{(N-j-1)\sigma}{K\nu+(N-j-1)\sigma}G_{j+1}^*(s)\right), & K \le j < N. \end{cases} \tag{22}$$

We can write this set of equations in matrix notation as $MG^* = R$, where

$$M = \begin{pmatrix} \nu+(N-1)\sigma+s & -(N-1)\sigma & & & & \\ -\nu & 2\nu+(N-2)\sigma+s & -(N-2)\sigma & & & \\ & -2\nu & 3\nu+(N-3)\sigma+s & -(N-3)\sigma & & \\ & \ddots & \ddots & \ddots & & \\ & & -\frac{N-2}{N-1}K\nu & K\nu+\sigma+s & -\sigma \\ & & & \frac{N-1}{N}K\nu & K\nu+s \end{pmatrix},$$

$G^* = \left(G_0^*(s), G_1^*(s), \ldots, G_{N-1}^*(s)\right)^T$ and $R = (\nu, \frac{2\nu}{2}, \ldots, \frac{K\nu}{K}, \frac{K\nu}{K+1}, \ldots, \frac{K\nu}{N})^T$. Borst, Boxma, and Hegde [1] also looked at this set of equations. They observed that $|m_{i,i}| > \sum_{j,j\neq i}|m_{i,j}|$ for $0 \le i < N$, where $m_{i,j}$ is the (i,j)th element of M. Matrices with this property are said to be strictly diagonally dominant. From matrix theory it can then be concluded that $\det(M) \neq 0$. This means that the Cramer's rule can be applied, thus $G_i^*(s) = \frac{\det(M_i)}{\det(M)}$, where M_i is identical to M with column i replaced by R. It is easy to see that $\det(M)$ is a polynomial in s of degree N and $\det(M_i)$ is a polynomial in s of degree $N-1$. It can be shown that the roots $\alpha_1, \ldots, \alpha_N$ of $\det(M)$ are real, unique and negative. We can thus use the partial fraction decomposition method and write $G_i^*(s) = \sum_{j=1}^N \frac{A_{i,j}}{s-\alpha_j}$ for certain $A_{i,j}$, which can be inverted to

$$P[G_i > t] = \sum_{j=1}^N \frac{A_{i,j}}{-\alpha_j} e^{\alpha_j t}, \tag{23}$$

and thus

$$P[D > t] = \sum_{i=1}^{N-1} r_i \sum_{j=1}^{N} \frac{A_{i,j}}{-\alpha_j} e^{\alpha_j t}. \tag{24}$$

3 Model of Wartenhorst

In Chapter 2 of [12] Wartenhorst also looks at a machine repair model, where there are N machines and K repairmen. The lifetime of every machine is $Exp(\sigma)$ distributed and the repair time is $Exp(\nu)$ distributed. The machines are repaired in FCFS order. The behavior of this model is described in Sections 2.2.1 and 2.3.1.

A special feature of his model is that the machines themselves also have jobs. Every machine has its own queue. We are interested in the behavior of the queue length at machine 1. At machine 1 jobs arrive according to a Poisson process with rate λ and the service times of these jobs are $Exp(\mu)$ distributed. The jobs of this machine are served in FCFS order. The N queues are independent of each other, jobs don't switch between queues. Notice that the arrival and service processes of other machines do not influence the behavior of the queue at machine 1, and thus are not specified.

The fraction of time machine 1 is available is given by:

$$P_{up} \equiv \frac{E[\text{lifetime}]}{E[\text{lifetime}] + E[D]} = \frac{1}{1 + \sigma E[D]}, \tag{25}$$

where D again is the length of an arbitrary downtime. The fraction of time that machine 1 is working, given the machine is up then is

$$\rho_{eff} \equiv \frac{\lambda}{\mu} \frac{1}{P_{up}}, \tag{26}$$

which is assumed to be less than 1.

Wartenhorst gives an exact solution of the steady state marginal queue length distribution. Thereafter he gives an approximation by assuming that the successive downtimes of a given machine are independent. To see that this generally is not the case, assume that a repair time of machine 1 is longer than normal. During this repair probably more machines break down than normal. So the next time machine 1 breaks down, more machines are in the queue than ordinarily is the case.

3.1 Exact solution

To give an exact solution of the steady state marginal queue length distribution he studies an $M/M/1$ queue in a Markovian environment S, that is the process $\{(Y(t), X(t)) | t \geq 0\}$, where $Y(t)$ is the number of jobs in the queue of machine 1 and the environment $X(t) \in S$ is defined by the triple (I, N_1, N_2),

$$I = \begin{cases} 0 & \text{if machine 1 is up,} \\ 1 & \text{if machine 1 is down,} \end{cases} \tag{27}$$

$$N_1 = \text{number of broken machines succeeding machine 1 in queue waiting for (or under) repair,} \tag{28}$$

$$N_2 = \text{number of broken machines preceding machine 1 in queue waiting for (or under) repair.} \tag{29}$$

Let the vector $\underline{\pi}$ be defined by

$$\pi_s = \lim_{t\to\infty} P[X(t) = s], \qquad s \in S. \tag{30}$$

These limiting probabilities exist, because the Markov Process $\{X(t)|t \geq 0\}$ has a finite number of states, which all communicate. To compute the joint probability distribution of the number of jobs in the queue of machine 1 and the environmental state, he defines vectors x_k by

$$x_{k,s} \equiv \lim_{t\to\infty} P(k \text{ jobs in queue at time } t \text{ and } X(t) = s), \tag{31}$$

where the job that is being processed is included.

This process is a Markov process, which has a matrix-geometric structure. It can be analyzed using the so-called matrix-geometric method that was developed by Neuts (see [7]). Using this method, he shows that the vectors $\underline{x}_k$ have the following form

$$\underline{x}_k = \underline{\pi}(I - R)R^k, \qquad k \geq 0. \tag{32}$$

The marginal distribution of the number of jobs at machine 1 is then given by

$$\underline{x}_k \cdot \underline{e}, \qquad k \geq 0, \tag{33}$$

where $\underline{e}$ is a vector with $|S|$ ones.

By conditioning on the event that the environmental state is in $A \subseteq S$, some interesting distributions can be derived. These conditional queue length distributions are given by

$$\lim_{t\to\infty} P[Y(t) = k|\text{the environmental state is in } A \subseteq S] = \frac{\sum_{s\in A} x_{k,s}}{\sum_{s\in A} \pi_j}. \tag{34}$$

By properly choosing A, distributions like that of the number of jobs in queue 1 at the beginning of an arbitrary up- or downperiod can be derived. This gives more insight in the build-up of the queue during downperiods than the marginal queue length distribution.

3.2 Approximation

As said above an approximation of the number of jobs in the queue of machine 1 is given by assuming the successive downtimes are independent and thus every downtime is distributed as D. This model with independent interrupts is known as the vacation model and has been extensively studied (see for instance Chapter 2 of [10]).

Wartenhorst examines the accuracy of this approximation. He concludes this approximation is exact if $K \geq N$, because there is no interference between the different machines. If the number of broken machines doesn't exceed the number of repairmen too often, this approximation still works pretty good. Another factor that influences the accuracy is the effective traffic intensity ρ_{eff}. If this intensity increases, severe queue build-up is more likely to occur and thus the interference between the downtimes affects the accuracy more. The largest influence however has the speed at which the jobs move through the system. If both λ and μ increase, but ρ stays the same, the approximation is less accurate. This can be explained by the fact that due to a small variation in the downperiod of the machine the number of jobs that have arrived during this downperiod also varies more.

4 Model of Mitrani et al.

In [6] and [11] Mitrani studies, together with Wright and Thomas respectively, the machine repair model from a different perspective. First of all he assumes that the machines break down and are repaired independently of each other. In his model there is just one arrival stream of jobs to be handled by the machines and jobs are routed to different machines with a certain probability, which may depend on which machines are up and down. In the first paper all jobs that are in the queue of a machine that breaks down are lost and all jobs that arrive during a downperiod are redirected to a different machine if any are available and else they too are lost. In the second paper he assumes that jobs aren't lost when a machine breaks down but remain in the queue of the machine until it is up again and the job can be handled.

The purpose of the analysis of these models is to compare different routing policies under different circumstances to minimize the loss of jobs and their sojourn times. A summary of both papers will be given in the next sections.

4.1 Routing with losing jobs

Consider N machines that are subject to breakdowns. Machine i is alternatingly up for an exponential time with mean $1/\xi_i$ and down for an exponential time with mean $1/\eta_i$. All up and down times are independent of each other. Let $\sigma \subseteq \{1, \ldots, N\}$ be the set of machines that is currently working and $\overline{\sigma}$ the set of machines that is currently down. Every machine has its own queue with infinite capacity.

Jobs arrive according to a Poisson process with rate λ. These jobs are then routed to queue i with probability $q_i(\sigma)$, i.e. these probabilities are allowed to depend on the set of machines that are currently working, but not on the different queue lengths. Jobs are never routed to machines that are down, when all machines are down the job gets lost. We thus have that:

$$\sum_{i \in \sigma} q_i(\sigma) = 1, \qquad \sigma \subseteq \{1, \ldots, N\}, \sigma \neq \emptyset. \tag{35}$$

When a machine breaks down all jobs in its queue, including the one in service, get lost. All service times are independent and $Exp(\mu_i)$ distributed. So the service times are allowed to depend on the machine that processes the job.

Note that we don't need stability conditions, since all queues are emptied once in a while.

To analyse this model, Mitrani introduces the Markov process $\{(I(t), J(t))|t \geq 0\}$, where $I(t) \subseteq \{1, \ldots, N\}$ is the set of machines that is working at time t and $J(t)$ is an integer vector of length $|I(t)|$ with the queue lengths of the working machines, including the jobs in service. Note that $J(t)$ is absent if $I(t) = \emptyset$. Let

$$p_\sigma(\mathbf{n}) = \lim_{t \to \infty} \mathbb{P}[I(t) = \sigma, J(t) = \mathbf{n}]. \tag{36}$$

Next, the balance equations for $p_\sigma(\mathbf{n})$ are given. Since the joint queue size distribution is very difficult to analyze, even for $N = 2$, Mitrani analyzes the marginal queue size distributions. To analyze the marginal queue size of queue i, the balance equations are summed over all $\mathbf{n}$ where $n_i = n$ for given configuration $i \in \sigma \subseteq \{1, \ldots, N\}$ and queue size n. Next these equations are multiplied by z^n and summed over all $n \geq 0$ to get equations for the generating functions of the marginal queue size distributions. Since there are 2^{N-1} configurations where machine i is working, we have a system of 2^{N-1} equations.

The problem is, however, that the generating functions are not the only unknowns in this system of equations, but this system also contains the 2^{N-1} unknown probabilities that queue i is empty for a given configuration. To overcome this problem, this system of equations is written in matrix form, where the unknown probabilities that queue i is empty are considered as constants. Then Cramer's rule is applied. Since the solutions that are attained are generating functions, they should be analytic inside the unit disc. Thus, when the denominator of Cramer's formula has a root inside the unit disc, also the numerator should be zero at that root, which gives extra equations for the unknown probabilities. Mitrani states that there should be exactly 2^{N-1} independent such equations in order for the balance equations to have a unique solution, although he wasn't able to prove that.

Now we know how to compute the generating functions of the marginal queue sizes, we can easily obtain various performance measures, such as the mean sojourn time and the mean number of jobs lost per unit of time. Numerical analysis by Mitrani for $N = 2$ shows that the results can be counter-intuitive. When the two machines work at the same speed, both work slower than the arrival rate, and break down and are repaired at the same rate, the jobs should be split equally among the two machines, when they are both working. But when one of the machines breaks down more often, more jobs should be routed to this machine! Mitrani explains this by noting that both queues tend to grow when the arrival rate at a certain queue is bigger than the speed of that machine. If it takes a long time before such a machine breaks down, the sojourn times are very long, and when it breaks down this causes many job losses. So it is beneficial to route more jobs to the less stable machine in order to keep the queue sizes relatively short.

4.2 Routing without losing jobs

In [11] the model as in Section 4.1 is studied with two differences:

- When a machine breaks down, the jobs in its queue including the one in service remain in place. Once the machine is up again, the service is resumed.
- Jobs may also be routed to machines that are down, but the routing probabilities may still depend on the system configuration. No jobs get lost in this model.

Since the queues aren't emptied anymore, we now have to require for stability that the average arrival rate of jobs that is routed to queue i is less than the service capacity of machine i.

Again, the Markov process $\{(I(t), J(t))|t \geq 0\}$ is studied, where $I(t) \subseteq \{1, \ldots, N\}$ is the set of machines that is working at time t and $J(t)$ now is an integer vector of length N with the queue lengths of the N machines, including the jobs in service. The analysis of the joint queue length distribution is very complicated, even for $N = 2$. However, to determine performance measures such as the mean sojourn time and the mean queue lengths it again suffices to look at the marginal queue size distributions. Since the arrivals and departures in a given queue only depend on the system configuration and not on the lengths of other queues, the process $\{(I(t), J_i(t))|t \geq 0\}$ also is a Markov process, where $J_i(t)$ is the length of queue i at time t, and thus the balance equations for this process can be formulated immediately.

This process is easier to analyze than the model of Section 4.1, since this process is a so called quasi-birth-and-death process, i.e. a transition is either a change of the configuration of the machines, or an arrival of a job, or a service completion of a job. Therefore, the spectral expansion method [5] can be used. This is not the case for the previous model, since there the queues can be instantly emptied, so an arbitrary amount of jobs can leave the system.

Let $\mathbf{p_i}(n)$ be the row vector with the equilibrium probabilities that queue i has n jobs in its queue for the different machine configurations. The balance equations for $n \geq 1$ can then be written in matrix form as:

$$\mathbf{p_i}(n-1)Q_{i,0} + \mathbf{p_i}(n)Q_{i,1} + \mathbf{p_i}(n+1)Q_{i,2} = 0, \qquad n = 1, 2, \ldots, \tag{37}$$

for certain matrices $Q_{i,0}, Q_{i,1}$ and $Q_{i,2}$, or equivalently as:

$$\mathbf{p_i}(n)Q_{i,0} + \mathbf{p_i}(n+1)Q_{i,1} + \mathbf{p_i}(n+2)Q_{i,2} = 0, \qquad n = 0, 1, \ldots. \tag{38}$$

Let

$$Q_i(z) := Q_{i,0} + Q_{i,1}z + Q_{i,2}z^2. \tag{39}$$

Any linear combination of the generalized eigenvectors of $Q_i(z)$ multiplied by the corresponding eigenvalue to the power n is a solution for (38). Here, $\mathbf{x}$ is generalized eigenvector of $Q_i(z)$ with corresponding eigenvalue α if

$$\mathbf{x}Q_i(\alpha) = 0. \tag{40}$$

The correct linear combination can determined by the balance equations for $n = 0$ and the normalization constraint. This procedure can always be performed when the ergodicity conditions hold.

Now it is known how to compute the marginal queue length distributions in the long run, and thus various performance measures can be computed, it can be investigated what routing policy should be chosen. Mitrani starts by assigning fixed weights to the machines in order to reduce the number of parameters that has to be set. Routing decisions are then only made by selecting a subset of machines where jobs can be routed to and then routes jobs based on the weights of these machines. Two types of routing policies are analyzed: one fixed strategy where the routing is done without considering the current machine configuration and one selective strategy where jobs are only routed to machines that are up, unless that is not possible. It turns out that in most cases the latter works much better than the first as was to be expected. There can be found examples however, where the fixed strategy outperforms the selective one.

The optimization of the weights is complex, since it involves a search in an N-dimensional space. Choosing these weights, however, is very important since it significantly influences the performance. Mitrani suggests to assign as weights the service capacities of the machines. This turns out to be a pretty good heuristic in practice.

5 Unreliable PS-machine with finite capacity

We will now consider an unreliable machine that serves jobs according to the Processor Sharing discipline, with a capacity of N jobs. These jobs arrive according to a Poisson process with rate λ. If an arriving job finds N jobs at the machine, the job is lost. The service times of jobs are i.i.d. and $Exp(\mu)$ distributed.

We assume the lifetimes of the machine are i.i.d. and are $Exp(\sigma)$ distributed. The repair times are independent of the lifetimes and they themselves are i.i.d. and $Exp(\nu)$ distributed. When the machine breaks down, all jobs will remain in the system and new jobs will still be accepted as long as the arriving jobs find less than N jobs at the machine.

5.1 Equilibrium distribution

To determine the equilibrium distribution we introduce the Markov process $\{(I(t), J(t))|t \geq 0\}$, where $I(t) = 0$ if the machine is down at time t and $I(t) = 1$ if the machine is up at time t. $J(t)$ is the number of jobs in the system at time t. Let

$$p_i(n) = \lim_{t\to\infty} \mathbb{P}[I(t) = i, J(t) = n], \qquad i = 0, 1, j = 0, \ldots, N. \tag{41}$$

Since we want to determine the equilibrium distribution using the spectral expansion method ([5]) we let $\mathbf{p}(n) = (p_0(n), p_1(n))$ and write the balance equations in matrix notation:

$$\mathbf{p}(0)\begin{pmatrix} \lambda+\nu & 0 \\ 0 & \lambda+\sigma \end{pmatrix} = \mathbf{p}(0)\begin{pmatrix} 0 & \nu \\ \sigma & 0 \end{pmatrix} + \mathbf{p}(1)\begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}, \tag{42}$$

$$\begin{aligned} \mathbf{p}(n)\begin{pmatrix} \lambda+\nu & 0 \\ 0 & \lambda+\mu+\sigma \end{pmatrix} &= \mathbf{p}(n-1)\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} + \mathbf{p}(n)\begin{pmatrix} 0 & \nu \\ \sigma & 0 \end{pmatrix} \\ &\quad + \mathbf{p}(n+1)\begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}, \qquad n = 1, \ldots, N-1, \end{aligned} \tag{43}$$

$$\mathbf{p}(N)\begin{pmatrix} \nu & 0 \\ 0 & \mu+\sigma \end{pmatrix} = \mathbf{p}(N-1)\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} + \mathbf{p}(N)\begin{pmatrix} 0 & \nu \\ \sigma & 0 \end{pmatrix}. \tag{44}$$

We first want to solve equation (43) and rewrite this equation as:

$$\mathbf{p}(n)Q_0 + \mathbf{p}(n+1)Q_1 + \mathbf{p}(n+2)Q_2 = 0, \qquad n = 0, \ldots, N-2, \tag{45}$$

with

$$Q_0 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad Q_1 = \begin{pmatrix} -(\lambda+\nu) & \nu \\ \sigma & -(\lambda+\mu+\sigma) \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}. \tag{46}$$

Let $Q(z) = Q_0 + Q_1 z + Q_2 z^2$, where $Q_i z$ means that all elements of Q_i have to be multiplied by z. We are now interested in the generalized left eigenvalues and eigenvectors of $Q(z)$, that is those vectors $\mathbf{x}_i$ and corresponding values z_i for which $\mathbf{x}_i Q(z_i) = 0$. Since $deg(det(Q(z))) = 3$, there are 3 such eigenvalues. Then, for arbitrary constants ξ_i, the vectors:

$$\mathbf{p}(j) = \sum_{i=1}^{3} \xi_i \mathbf{x}_i z_i^j, \qquad j = 0, \dots N, \tag{47}$$

are a solution of (45).

To find the eigenvalues, we have to solve:

$$\begin{aligned} 0 &= det(Q(z)) \\ &= (\lambda - (\lambda+\nu)z)(\lambda - (\lambda+\mu+\sigma)z + \mu z^2) - \sigma\nu z^2 \\ &= -(\lambda\mu + \mu\nu)(z-1)\left(z^2 - \frac{\lambda^2 + \lambda\mu + \lambda\nu + \lambda\sigma}{\lambda\mu + \mu\nu} z + \frac{\lambda^2}{\lambda\mu + \mu\nu}\right) \end{aligned} \tag{48}$$

The discriminant of the quadratic part of this equation is:

$$\left(\frac{\lambda^2 + \lambda\mu + \lambda\nu + \lambda\sigma}{\lambda\mu + \mu\nu}\right)^2 - 4\frac{\lambda^2}{\lambda\mu + \mu\nu} = \frac{\lambda^2((\lambda - \mu + \nu)^2 + 2(\lambda + \mu + \nu)\sigma + \sigma^2)}{(\lambda\mu + \mu\nu)^2}, \tag{49}$$

and since all coefficients are strictly positive, the discriminant is also strictly positive and thus there are two different real roots of this quadratic equation. So the three eigenvalues are:

$$z_1 = 1, \tag{50}$$

$$z_2 = \frac{\lambda^2 + \lambda\mu + \lambda\nu + \lambda\sigma + \lambda\sqrt{(\lambda - \mu + \nu)^2 + 2(\lambda + \mu + \nu)\sigma + \sigma^2}}{2(\lambda\mu + \mu\nu)}, \tag{51}$$

$$z_3 = \frac{\lambda^2 + \lambda\mu + \lambda\nu + \lambda\sigma - \lambda\sqrt{(\lambda - \mu + \nu)^2 + 2(\lambda + \mu + \nu)\sigma + \sigma^2}}{2(\lambda\mu + \mu\nu)}. \tag{52}$$

It can easily be verified that corresponding eigenvalues are:

$$\mathbf{x}_1 = (\sigma, \nu), \tag{53}$$

$$\mathbf{x}_2 = (\sigma, \nu + \lambda - \frac{\lambda}{z_2}), \tag{54}$$

$$\mathbf{x}_3 = (\sigma, \nu + \lambda - \frac{\lambda}{z_3}). \tag{55}$$

Assume that $z_2, z_3 \neq 1$, which will almost always be the case in practice. Then these three eigenvectors are linearly independent, and thus span a 3-dimensional space. When N would be infinite, the solution space of (45) is linear and its dimensionality is exactly 3 ([3]), so every general solution of (45) can be written as in (47). Notice however, that if we add an arbitrary constant, ξ_4 say, to $p_0(N)$, then equation (45) will still hold. The coefficients $\xi_i, i = 1, \dots, 4$ can now be determined from equations (42) and (44) and the normalization constraint.

5.2 Sojourn time distribution

Next, we are going to look at the sojourn time distribution S of an arbitrary job. We are going to do this in a similar way as in Section 2.3. So we condition on the state the system is in found by a job that just arrived (the tagged job). By the PASTA property ([13]), the state distribution at an arrival epoch is equal to the state distribution in equilibrium, which we computed in the previous section.

When the tagged job finds N jobs in the system, he immediately leaves the system. When the tagged job finds j jobs in the system, $j < N$, and the machine is down, an event happens after an $Exp(\lambda + \nu)$ amount of time. With probability $\frac{\lambda}{\lambda+\nu}$ this event is an arrival of a new job and it is just like the tagged job arrives and finds $\min\{j+1, N-1\}$ jobs in the system and the machine is still down. With probability $\frac{\nu}{\lambda+\nu}$ the machine has been repaired and it is just like the tagged job arrives and finds j jobs in the system and the machine is working.

When the tagged job finds j jobs in the system, $j < N$, and the machine is working, an event happens after an $Exp(\lambda+\mu+\sigma)$ amount of time. With probability $\frac{\lambda}{\lambda+\mu+\sigma}$ a new job arrives and it is just like the tagged job arrives and finds $\min\{j+1, N-1\}$ jobs in the system and the machine is still working. With probability $\frac{\mu}{\lambda+\mu+\sigma}$ a job has been served. With probability $\frac{1}{j+1}$, this job is the tagged job and this job leaves the system, with probability $\frac{j}{j+1}$ an other job has been served and it is just like the tagged job arrives and finds $j-1$ jobs in the system and the machine is still working. Finally, with probability $\frac{\sigma}{\lambda+\mu+\sigma}$ the machines breaks down and it is just like the tagged job arrives, finds j jobs in the system and the machine is down.

Define $G_{0,j}$ as the sojourn time of a random job, given that he finds a broken machine and j jobs in the system and define $G_{1,j}$ as the sojourn time of a random job, given that he finds a working machine and j jobs in the system. Let $G^*_{i,j}(s)$ be the Laplace-Stieltjes transform of $G_{i,j}$. Then we have the following recurrence relation for $G^*_{i,j}(s)$:

$$G^*_{i,j}(s) = \begin{cases} 1, & i=0,1, j=N, \\ \frac{\lambda+\nu}{\lambda+\nu+s}\left(\frac{\lambda}{\lambda+\nu}G^*_{0,\min\{j+1,N-1\}}(s)\right. & \\ \qquad \left. + \frac{\nu}{\lambda+\nu}G^*_{1,j}(s)\right), & i=0, j<N, \\ \frac{\lambda+\mu+\sigma}{\lambda+\mu+\sigma+s}\left(\frac{\lambda}{\lambda+\mu+\sigma}G^*_{1,\min\{j+1,N-1\}}(s)\right. & \\ \qquad + \frac{\mu}{\lambda+\mu+\sigma}\left(\frac{1}{j+1}1 + \frac{j}{j+1}G^*_{1,j-1}(s)\right) & \\ \qquad \left. + \frac{\sigma}{\lambda+\mu+\sigma}G^*_{0,j}(s)\right), & i=1, j<N. \end{cases} \tag{56}$$

We can write this set of equations in matrix notation as $MG^* = R$, where

$$M = \left(\begin{array}{cccc|cccc} \lambda+\nu+s & -\lambda & & & -\nu & & & \\ & \ddots & \ddots & & & \ddots & & \\ & & \lambda+\nu+s & -\lambda & & & -\nu & \\ & & & \nu+s & & & & -\nu \\ \hline -\sigma & & & & \lambda+\mu+\sigma+s & -\lambda & & \\ & -\sigma & & & -\frac{1}{2}\mu & \lambda+\mu+\sigma+s & -\lambda & \\ & & \ddots & & & \ddots & \ddots & \ddots \\ & & -\sigma & & & -\frac{N-2}{N-1}\mu & \lambda+\mu+\sigma+s & -\lambda \\ & & & -\sigma & & & -\frac{N-1}{N}\mu & \mu+\sigma+s \end{array}\right),$$

$G^* = \left(G^*_{0,0}(s), \ldots, G^*_{0,N-1}(s), G^*_{1,0}(s), \ldots, G^*_{1,N-1}(s)\right)^T$ and $R = \left(0, \ldots, 0, \mu, \frac{1}{2}\mu, \ldots, \frac{1}{N-1}\mu, \frac{1}{N}\mu\right)^T$. The matrix M is strictly diagonally dominant, so $\det(M) \neq 0$. This means that the Cramer's rule can be applied, thus $G^*_i(s) = \frac{\det(M_i)}{\det(M)}$, where M_i is identical to M with column i replaced by R. Because the degree of the numerator is smaller than the degree of the denominator, we can do a fraction decomposition, which can be easily inverted. This may again lead to a phase-type distribution.

5.3 Comparison with FCFS

We are now going to compare the above results with a similar machine, but now assume that jobs are being processed according to the First Come, First Serve discipline instead of the Processor Sharing discipline. The equilibrium distribution is of course the same as with PS, since the balance equations are the same.

The sojourn time of an arbitrary job can be determined more easily then with PS, because jobs that arrive after the tagged job do not influence the sojourn time of the tagged job. By similar

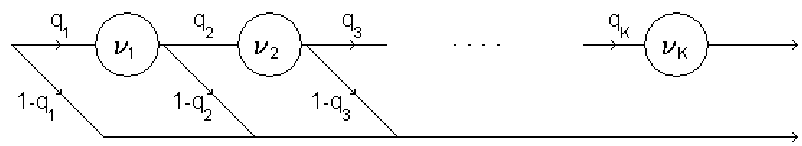

Figure 4: When the downtime can be represented as a K-phase Coxian distribution, with probability $1-q_1$ the downtime is 0 and with probability q_1 the machine remains down for an $Exp(\nu_1)$ amount of time. Then with probability $1-q_2$ the machine goes up again, etc.

arguments as above, we have the following recurrence relation for $G^*_{i,j}(s)$:

$$G^*_{i,j}(s) = \begin{cases} 1, & i=0,1, j=N, \\ \frac{\nu}{\nu+s}G^*_{1,j}(s), & i=0, j<N, \\ \frac{\mu+\sigma}{\mu+\sigma+s}\left(\frac{\mu}{\mu+\sigma}G^*_{1,j-1}(s)+\frac{\sigma}{\mu+\sigma}G^*_{0,j}(s)\right), & i=1, j<N, \end{cases} \tag{57}$$

where $G^*_{1,-1}(s)=1$ by definition. We thus have, for $j<N$, that:

$$\begin{aligned}(\mu+\sigma+s)G^*_{1,j}(s) &= \mu G^*_{1,j-1}(s)+\sigma G^*_{0,j}(s) \\ &= \mu G^*_{1,j-1}(s)+\frac{\sigma\nu}{\nu+s}G^*_{1,j}(s).\end{aligned} \tag{58}$$

So

$$\begin{aligned}G^*_{1,j}(s) &= \frac{\mu(\nu+s)}{(\mu+\sigma+s)(\nu+s)-\sigma\nu}G^*_{1,j-1}(s) \\ &= \frac{\mu\nu+\mu s}{\mu\nu+(\mu+\nu+\sigma)s+s^2}G^*_{1,j-1}(s) \\ &= \left(\frac{\mu\nu+\mu s}{\mu\nu+(\mu+\nu+\sigma)s+s^2}\right)^{j+1} \\ &= \left(\frac{\mu}{\mu+s}\frac{(\mu+s)(\nu+s)}{(\mu+s)(\nu+s)+s\sigma}\right)^{j+1}.\end{aligned} \tag{59}$$

It turns out that this is exactly the Laplace-Stieltjes transform of $j+1$ times a service time plus a geometrically distributed number of downtimes.

5.4 General downtime distribution

When the downtime of the machine has a general distribution that can be well-represented by a Coxian distribution ([8]), an analysis as above is also very useful. These downtimes can for instance be seen as a busy period of high-priority jobs that arrive with rate σ.

Suppose the downtime of the machine has a K-phase Coxian distribution with parameters $q_1,\dots,q_K,\nu_1,\dots,\nu_K$ (see Figure 4). To analyze the equilibrium distribution of this system, we not only have to store the number of jobs in the system and whether the system is up or down, but when the machine is down also the downphase the machine is in. Since a fraction of $1-q_1$ of the times the breakdown of the machine is a false one, i.e. the downtime is zero, the rate at which the system jumps to the first downphase is equal to $q_1\sigma$. The balance equations can now be posed again and written in matrix notation as above. This system can again be analyzed with the spectral expansion method. It should be noted however, that the dimension of the matrix is larger than two, and thus the eigenvalues cannot be determined analytically anymore.

To compute the sojourn time of an arbitrary job we need the Laplace-Stieltjes transforms $D^*_k(s)$ of the remaining downtime when the machine is in downphase k. These can be recursively computed using

$$D_k^*(s) = \frac{\nu_k}{\nu_k + s}((1 - q_{k+1}) \cdot 1 + q_{k+1} D_{k+1}^*(s)), \qquad k = 1, \ldots, K, \tag{60}$$

where $q_{K+1} = 0$ by definition. The rest of the analysis can be done as above, both for PS and FCFS.

6 Conclusion

In this report we discussed several machine repair models. These models may help to analyze real life situations and analyze various performance measures, such as the number of machines that is down, the downtime distribution and the effect that this may have on the jobs that are being served by unreliable machines. It also helps to find a good routing policy when jobs may be handled by different machines.

Two techniques especially have helped in doing the analysis. The first technique is the so-called spectral expansion method ([5]), which is a better technique to use to analyze the equilibrium distribution of the number of jobs in a queue than the matrix-geometric method ([7]). This technique however can only be used when the number of states the machine is in (excluding the number of jobs in its queue) is finite and the number of jobs that can enter or leave the system at once is limited.

The second technique that is used is using Laplace-Stieltjes transform of the remaining downtime or sojourn time given the current state. Because of the memoryless properties that are assumed most of the time, we can then recursively determine the downtimes or sojourn times given the state of the system at the moment a machine breaks down or a job enters the system. Combining this with the equilibrium distribution of the state of the system and because of the PASTA property ([13]), we can then determine the general arbitrary downtime or sojourn time distribution.

References

[1] S.C. Borst, O.J. Boxma and N. Hegde, *Sojourn Times in Finite-Capacity Processor-Sharing Queues.* In: R. Sabella (ed.), *Proc. Euro-NGI Conference, Rome*, 2005.

[2] J.W. Cohen, *On Regenerative Processes in Queueing Theory*, Springer-Verlag, 1976.

[3] I. Gohberg, P. Lancaster and L. Rodman, *Matrix Polynomials*, p. 327, Academic Press, 1982.

[4] S.S. Lavenberg and M. Reiser, *Stationary State Probabilities at Arrival Instants for Closed Queueing Networks with Multiple Types of Customers*, Journal of Applied Probability, vol 17, pp. 1048-1061, 1980.

[5] I. Mitrani and D. Mitra, *A Spectral Expansion Method for Random Walks on Semi-Infinite Strips.* In: R. Beauwens and P. de Groen (eds.), *Iterative Methods in Linear Algebra*, pp. 141-149, North-Holland, 1992.

[6] I. Mitrani and P.E. Wright, *Routing in the Presence of Breakdowns*, Performance Evaluation, vol. 20, pp. 151-164, 1994.

[7] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*, The Johns Hopkins University Press, 1981.

[8] T. Osogami and M. Harchol-Balter, *Necessary and sufficient conditions for representing general distributions by Coxians*, In: P. Kemper and W.H. Sanders (eds.), *TOOLS 2003*, LNCS 2794, pp. 182-199, Springer Berlin, 2003.

[9] R. Schassberger, *Warteschlangen*, Springer-Verlag, 1973.

[10] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1*, North-Holland, 1991.

[11] N. Thomas and I. Mitrani, *Routing Among Different Nodes Where Servers Break Down Without Losing Jobs.* In: F. Baccelli, A. Jean-Marie and I. Mitrani (eds.), *Quantitative Methods in Parallel Systems*, pp. 248-261, Springer-Verlag, 1995.

[12] P. Wartenhorst, *Performance Analysis of Repairable Systems*, Ph.D Thesis, Tilburg University, 1993.

[13] R.W. Wolff, *Poisson Arrivals See Time Averages*, Operations Research, Vol. 30, pp. 223-231, 1982.