Assigning identical operators to different machines

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 93-31

Assigning identical operators to different machines

A.G. de Kok
J. van der Wal

Eindhoven, November 1993
The Netherlands

Eindhoven University of Technology
Department of Mathematics and Computing Science
Probability theory, statistics, operations research and systems theory
P.O. Box 513
5600 MB Eindhoven - The Netherlands

Secretariat: Dommel building 0.03
Telephone: 040-47 3130

ISSN 0926 4493

ASSIGNING IDENTICAL OPERATORS TO DIFFERENT MACHINES

Jan van der Wal and Ton G. de Kok, Eindhoven University of Technology

ABSTRACT

This paper presents a new approximation for the response times in a production system with M machines and s operators, where each machine has its own arrival stream of jobs, where no machine can work without the constant attention of an operator and where the operators can serve only one machine at a time. The arrival streams are Poisson with equal rates and all jobs have the same service time distribution. Several strategies are considered for the operators. Extensive simulations show that for the case of exponential service times the differences between these strategies are usually small. The approximation is also quite good for non-exponential service times.

1. INTRODUCTION

We will consider a production system with M machines attended by s operators with $s < M$, so the number of operators is smaller than the number of machines. Machine m receives a Poisson arrival stream of jobs with rate λ_m. The streams for different machines are independent. No machine can work without the constant attention of an operator. An operator can serve only one machine at a time. All machines are different, so a job can only be realized on the machine where it arrived.

When controlling such a system the main problem is to avoid as much as possible states in which jobs are waiting while operators are idle. Clearly this may always happen as a job can arrive at a machine that is already busy when one or more operators are idle. So in this system the operators will be less efficient than in an $M \mid G \mid s$ system.

We will restrict our attention to the symmetric case with identical arrival rates, $\lambda_m = \lambda$ for all m, and the same service time distribution on all machines. For the case of exponential service times Durlinger and Geurts [1] give an approximation based on aggregation which is excellent if an operator who has completed a job randomly chooses an idle and waiting machine.

In this paper we present another approximation based on the asymptotic behaviour of the system which suggests a straightforward extension to the case of non-exponential service times. Further we study a number of strategies for the operator problem: which job to serve next. The differences are mostly small. The asymptotic approximation turns out to be very good for both low and high utilizations and practically all interesting strategies.

Throughout this paper the mean service time is taken to be 1 and the service times are assumed to be exponential except in section 7.

Section 2 introduces the six operator strategies that we will compare. In section 3 the approximation of Durlinger and Geurts is given. Section 4 contains the asymptotic approximation. Section 5 presents the results from the simulations. In section 6 we discuss the ergodicity conditions for the Shortest Queue strategy. Section 7 gives some results for non-exponential service times.

2 THE OPERATOR STRATEGIES

Whenever a job is completed an operator is "freed". Then it has to be decided where he should go to next.

As we stated before, this system will always be less efficient than the $M \mid G \mid s$ system and it is important to avoid as much as possible unbalanced states. (A state is called unbalanced if the difference in queue lenghts is large.) In an unbalanced state it is more likely that in the near future a state will be reached where jobs are waiting while one or more operators are idle. So the better strategies are the ones that try to balance the queue lengths.

In this paper we will compare six strategies.

1. Random Machine (RM). The operator selects at random one of the idle and waiting machines. All machines have equal weights.
2. Random Job (RJ). The operator selects at random one of the idle and waiting machines. The machines have weights equal to the number of jobs waiting. So the operators give some preference to the longer queues.
3. Longest Queue (LQ). The operator goes to (one of) the longest queue(s). One may show that this strategy is optimal, see [3].
4. First Come First Served (FCFS). The operator goes to the job which arrived first (considering only the jobs waiting at idle machines).
5. Longest Queue Exhaustive Service (LQES). An operator always completes all jobs at a machine before he becomes free. Then he goes, as in LQ, to the longest queue.

For comparison we also consider

6. Shortest Queue (SQ). After completing a job the operator goes to the idle and waiting machine with the shortest queue. This strategy maximizes the average response time, see [3].

If we denote the mean waiting time (response time) for strategy RM by W_{RM} (S_{RM}), etc., then we expect the following partial ordering of the waiting times:

$$W_{LQ} < W_{RJ} < W_{RM} < W_{SQ} ,$$

$$W_{LQ} < W_{FCFS} < W_{SQ} ,$$

and

$$W_{LQ} < W_{LQES} < W_{SQ} .$$

The inequality $W_{RJ} < W_{RM}$ is intuitively clear. Giving preference to the longer queue in some random way should help. The order for the strategies FCFS and LQES and their position relative to RJ and RM is unclear. Further one may expect that the differences are small for low loads but they might become considerable for high loads. The results from the simulations in section 5 will confirm this.

For all these strategies (except FCFS) the state of the system in the completely symmetric case is fully characterized by the state vector $(i_1, \cdots, i_M ; n_1, \cdots, n_M)$, with $i_m = 1$ if machine m is busy and $i_m = 0$ if machine m is idle, and n_m the number of jobs waiting (not in service!) at machine m. For none of these systems we have been able to obtain a simple expresssion for the equilibrium distribution.

3. AGGREGATION BASED ON THE BOSE-EINSTEIN ASSUMPTION

The approximation of Durlinger and Geurts [1] for the completely symmetric exponential case is based on the assumption that the system behaves almost according to the Bose-Einstein rule if an operator who just finished his job picks the next machine at random. Their approach is as follows.

Assume that the margimal distribution of the jobs at the various queues given that the total number of jobs in the system equals k follows the Bose-Einstein probability distribution. I.e., all possible distributions of k jobs over M queues are equally likely, so the probability $p(k_1, ..., k_M ; k)$ of finding k_m jobs in queue m, $m = 1, ..., M$, satisfies

$$p(k_1, ..., k_M ; k) = 1 \ / \ \binom{k + M - 1}{M - 1} \tag{1}$$

Remark

Note that this result is exact if the number of operators is equal to the number of machines, in which case each machine has its own operator and thus all machines are independent. Then, as one easily sees, the probability of finding microstate $(k_1, ..., k_M)$ is equal to $(1 - \rho)^M \rho^k$ with $k = \Sigma k_m$.

This assumption (for $s < M$) makes it possible to aggregate the microstates into macrostates. The macrostates are characterized by the total number of jobs waiting and in service in the system.

For this system the rate $\gamma(k-1, k)$ by which transitions from $k-1$ to k are made equals λM, the total arrival rate. The rate by which transitions are made from k to $k-1$ is equal to the expected number of active machines. This number is approximated using equation (1). The probability $p(l\,;k)$ that l out of the M queues are empty is given by

$$p(l\,;k) = \binom{M}{l}\binom{k-1}{M-l-1} \Big/ \binom{k+M-1}{M-1} \qquad (2)$$

If l queues are empty $M-l$ are nonempty and then the rate from k tot $k-1$ is equal to $min(s,\ M-l)$. So the mean rate $\gamma(k,k-1)$ for transitions from k to $k-1$ is given by

$$\gamma(k,k-1) = \sum_{l=0}^{M-1} p(l\,;k) min(s,\ M-l) \qquad (3)$$

This birth and death process for the macrostates is easily solved numerically.

In the sequel we call this approximation AGGR(egation).

4. AN APPROXIMATION BASED ON THE ASYMPTOTIC BEHAVIOUR

In this section we derive an approximation for the mean response time based on the asymptotic behaviour of the system. First, let us think a little about the asymptotic behaviour of the system if M becomes large. If the number of operators s and their utilization ρ remain constant while M tends to infinity then the arrival rate per machine $\lambda = s\rho/M$ tends to zero and the total number of jobs in the system converges in distribution to the number in the $M\,|M\,|s$ queue with arrival rate $s\rho$ and mean service time 1. Further, if $s = M$ then each queue behaves as an $M\,|M\,|1$ queue with arrival rate ρ and mean service time 1.

These considerations suggest that a convex combination of the response times in the $M\,|M\,|1$ and $M\,|M\,|s$ queues might work well. So we suggest an approximation for the mean waiting time $W(s,M,\rho)$ of the following form:

$$W(s,M,\rho) = \alpha(s,M,\rho)\ W_1(\rho) + (1-\alpha(s,M,\rho))\ W_s(\rho)\ , \qquad (4)$$

where $W_c(\rho)$ is the mean waiting time in an $M\,|M\,|c$ queue with mean service time 1. The problem is to find a suitable function $\alpha(s,M,\rho)$.

A primitive attempt is to plot the ideal factor α for a number of simulation results. We started with a testbed of 88 simulations with M equal to 4, 8 and 16 with s running from 2 to $M-1$ and utilizations 0.7, 0.8, 0.9 and 0.95 for which the strategies RM and RJ were simulated. For this set we found excellent results for a function of the form

$$\alpha(s,M,\rho) = (s/M)^{\frac{1}{p(1-\rho)+q(1-s/M)+r}}\ , \qquad (5)$$

with approximately $p = 0.88$, $q = 0.35$ and $r = 0.02$.

This function yields a maximum absolute relative error in the mean response time for this testbed of 10 percent and an average absolute error of less than 2 percent. However we have no understanding for this function at all and when we extended the testbed with the 44 simulations for the utilizations 0.3 and 0.5 the quality of the approximation dropped considerably.

Being not very pleased with this function we looked for a more sophisticated way of obtaining α. Let us again think about what happens if M and s are large. If s is large, then for utilization ρ on the average $s(1-\rho)$ operators are free. So for large s there will "always" be a free operator. Now look at one machine. For that machine the arrival rate is equal to $s\rho/M$. As there is always an operator available this queue will behave (approximately) as an $M\,|M\,|\,1$ with arrival rate $s\rho/M$ and mean service time 1. So the mean waiting time for each machine is approximately equal to $W_1(s\rho/M) = (s\rho/M)\,/\,(1 - s\rho/M)$. Further $W_s(\rho)$ goes to 0 if s tends to infinity, so for large s we should have

$$\alpha(s,M,\rho) = \frac{W_1(s\rho/M)}{W_1(\rho)}. \tag{6}$$

The approximation (4) with α given by (6), further denoted by ASYMP, will turn out to be very good, even for $\rho = 0.95$.

5. SIMULATION RESULTS

We have run extensive simulations for the testbed of 132 cases: M equal to 4, 8 and 16, s running from 2 to $M-1$ and operator utilizations 0.3, 0.5, 0.7, 0.8, 0.9 and 0.95.
Tabels 5.1 to 5.6 give for each of the six strategies the results per utilization. In all simulations the standard deviation of the point estimate for S is around 2 promille of the mean. The simulations for utilization 0.95 require per case and per strategy already a couple of hours on a Sun Sparc workstation.

Response times for ρ = 0.30									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	1.227	1.225	1.225	1.224	1.226	1.225	1.231	1.235
4	**3**	1.297	1.296	1.296	1.296	1.297	1.297	1.298	1.301
8	**2**	1.156	1.155	1.154	1.154	1.155	1.155	1.158	1.161
8	**3**	1.145	1.144	1.144	1.145	1.145	1.145	1.147	1.150
8	**4**	1.181	1.181	1.181	1.180	1.181	1.181	1.182	1.184
8	**5**	1.231	1.233	1.233	1.232	1.231	1.231	1.232	1.233
8	**6**	1.291	1.291	1.291	1.289	1.291	1.291	1.291	1.291
8	**7**	1.357	1.356	1.356	1.356	1.357	1.357	1.356	1.356
16	**2**	1.126	1.125	1.125	1.123	1.125	1.125	1.126	1.129
16	**3**	1.085	1.085	1.085	1.086	1.085	1.085	1.086	1.088
16	**4**	1.089	1.089	1.089	1.090	1.089	1.089	1.090	1.092
16	**5**	1.106	1.106	1.106	1.107	1.106	1.106	1.107	1.108
16	**6**	1.128	1.128	1.128	1.128	1.128	1.128	1.128	1.129
16	**7**	1.151	1.151	1.151	1.152	1.151	1.151	1.151	1.152
16	**8**	1.177	1.177	1.177	1.178	1.177	1.177	1.177	1.177
16	**9**	1.204	1.204	1.204	1.203	1.204	1.204	1.203	1.203
16	**10**	1.231	1.231	1.231	1.231	1.231	1.231	1.231	1.231
16	**11**	1.259	1.260	1.260	1.259	1.259	1.259	1.260	1.260
16	**12**	1.290	1.290	1.290	1.290	1.290	1.290	1.290	1.290
16	**13**	1.324	1.323	1.323	1.325	1.324	1.324	1.322	1.322
16	**14**	1.356	1.356	1.356	1.356	1.356	1.356	1.356	1.356
16	**15**	1.389	1.391	1.391	1.389	1.389	1.389	1.391	1.391

Tabel 5.1

Response times for ρ = 0.50									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	1.544	1.534	1.530	1.535	1.531	1.528	1.552	1.556
4	**3**	1.650	1.644	1.643	1.643	1.643	1.643	1.658	1.663
8	**2**	1.421	1.417	1.415	1.415	1.414	1.412	1.422	1.429
8	**3**	1.342	1.335	1.335	1.334	1.336	1.335	1.344	1.352
8	**4**	1.379	1.377	1.375	1.374	1.375	1.374	1.383	1.391
8	**5**	1.474	1.470	1.469	1.469	1.471	1.472	1.476	1.483
8	**6**	1.601	1.601	1.600	1.604	1.599	1.600	1.608	1.613
8	**7**	1.777	1.775	1.775	1.775	1.778	1.776	1.780	1.783
16	**2**	1.373	1.373	1.372	1.368	1.368	1.368	1.373	1.378
16	**3**	1.239	1.237	1.236	1.234	1.236	1.235	1.239	1.245
16	**4**	1.209	1.208	1.208	1.209	1.207	1.206	1.211	1.217
16	**5**	1.222	1.218	1.218	1.219	1.220	1.218	1.222	1.228
16	**6**	1.249	1.248	1.248	1.249	1.248	1.248	1.251	1.256
16	**7**	1.291	1.288	1.288	1.289	1.290	1.290	1.291	1.296
16	**8**	1.339	1.337	1.337	1.339	1.339	1.338	1.339	1.343
16	**9**	1.393	1.394	1.394	1.392	1.392	1.392	1.395	1.398
16	**10**	1.452	1.454	1.454	1.455	1.453	1.454	1.456	1.458
16	**11**	1.523	1.520	1.519	1.523	1.524	1.523	1.525	1.526
16	**12**	1.597	1.598	1.599	1.596	1.597	1.597	1.600	1.601
16	**13**	1.682	1.682	1.682	1.683	1.682	1.682	1.684	1.685
16	**14**	1.777	1.775	1.776	1.774	1.778	1.778	1.778	1.778
16	**15**	1.880	1.877	1.877	1.882	1.882	1.882	1.882	1.883

Tabel 5.2

Response times for ρ = 0.70									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	2.323	2.250	2.233	2.242	2.230	2.211	2.293	2.278
4	**3**	2.431	2.370	2.344	2.352	2.333	2.330	2.423	2.393
8	**2**	2.102	2.075	2.068	2.064	2.060	2.050	2.077	2.086
8	**3**	1.828	1.791	1.781	1.784	1.782	1.776	1.811	1.820
8	**4**	1.813	1.783	1.775	1.773	1.769	1.765	1.806	1.813
8	**5**	1.946	1.910	1.904	1.906	1.902	1.897	1.939	1.946
8	**6**	2.197	2.171	2.158	2.166	2.162	2.161	2.197	2.204
8	**7**	2.609	2.607	2.588	2.600	2.595	2.606	2.620	2.627
16	**2**	2.021	2.011	2.005	2.007	2.005	2.005	2.010	2.017
16	**3**	1.660	1.651	1.648	1.646	1.643	1.640	1.652	1.663
16	**4**	1.532	1.521	1.517	1.515	1.515	1.512	1.524	1.537
16	**5**	1.494	1.478	1.477	1.477	1.473	1.474	1.489	1.502
16	**6**	1.504	1.494	1.489	1.490	1.484	1.485	1.501	1.514
16	**7**	1.546	1.537	1.529	1.530	1.528	1.527	1.545	1.558
16	**8**	1.613	1.601	1.600	1.597	1.600	1.597	1.613	1.625
16	**9**	1.698	1.693	1.691	1.687	1.686	1.688	1.703	1.715
16	**10**	1.814	1.808	1.802	1.802	1.801	1.803	1.815	1.827
16	**11**	1.950	1.944	1.945	1.944	1.940	1.941	1.953	1.965
16	**12**	2.118	2.116	2.113	2.109	2.108	2.111	2.121	2.132
16	**13**	2.320	2.325	2.317	2.319	2.315	2.313	2.328	2.338
16	**14**	2.570	2.571	2.578	2.580	2.571	2.574	2.585	2.592
16	**15**	2.906	2.897	2.903	2.897	2.902	2.891	2.910	2.915

Tabel 5.3

Response times for ρ = 0.80									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	3.369	3.148	3.090	3.099	3.090	3.073	3.196	3.148
4	**3**	3.452	3.174	3.107	3.147	3.088	3.068	3.297	3.174
8	**2**	2.995	2.909	2.896	2.895	2.896	2.877	2.910	2.917
8	**3**	2.465	2.369	2.350	2.350	2.333	2.324	2.389	2.392
8	**4**	2.373	2.263	2.236	2.227	2.225	2.207	2.296	2.288
8	**5**	2.494	2.390	2.351	2.351	2.343	2.331	2.436	2.416
8	**6**	2.842	2.746	2.715	2.700	2.694	2.682	2.800	2.770
8	**7**	3.522	3.450	3.426	3.414	3.419	3.394	3.506	3.478
16	**2**	2.878	2.836	2.821	2.831	2.833	2.816	2.831	2.840
16	**3**	2.235	2.193	2.188	2.187	2.183	2.173	2.195	2.208
16	**4**	1.976	1.928	1.922	1.920	1.910	1.910	1.934	1.949
16	**5**	1.860	1.813	1.803	1.803	1.796	1.797	1.826	1.841
16	**6**	1.830	1.785	1.781	1.772	1.769	1.768	1.799	1.814
16	**7**	1.851	1.812	1.798	1.798	1.793	1.787	1.825	1.839
16	**8**	1.921	1.877	1.866	1.860	1.851	1.852	1.893	1.905
16	**9**	2.017	1.978	1.968	1.960	1.956	1.951	1.998	2.009
16	**10**	2.161	2.125	2.112	2.107	2.103	2.098	2.143	2.153
16	**11**	2.349	2.313	2.304	2.296	2.290	2.287	2.335	2.345
16	**12**	2.599	2.567	2.553	2.548	2.545	2.543	2.587	2.596
16	**13**	2.910	2.900	2.889	2.881	2.882	2.879	2.919	2.929
16	**14**	3.366	3.356	3.344	3.333	3.336	3.331	3.372	3.383
16	**15**	4.009	4.015	3.997	3.986	4.000	3.991	4.017	4.027

Tabel 5.4

Response times for ρ = 0.90									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	7.430	5.746	5.633	5.679	5.619	5.579	5.820	5.694
4	**3**	8.268	5.443	5.130	5.274	5.044	4.964	5.666	5.172
8	**2**	5.918	5.401	5.384	5.371	5.395	5.376	5.412	5.416
8	**3**	4.767	4.092	4.027	4.035	4.000	3.993	4.092	4.079
8	**4**	4.397	3.643	3.545	3.543	3.524	3.490	3.664	3.609
8	**5**	4.546	3.662	3.518	3.519	3.481	3.445	3.724	3.593
8	**6**	5.130	4.149	3.962	3.937	3.885	3.845	4.270	4.026
8	**7**	6.504	5.524	5.275	5.217	5.147	5.116	5.680	5.311
16	**2**	5.516	5.316	5.326	5.305	5.349	5.325	5.320	5.330
16	**3**	4.072	3.848	3.832	3.813	3.812	3.823	3.851	3.865
16	**4**	3.426	3.180	3.167	3.150	3.147	3.142	3.180	3.196
16	**5**	3.101	2.832	2.799	2.799	2.788	2.776	2.835	2.850
16	**6**	2.921	2.651	2.622	2.615	2.607	2.597	2.663	2.673
16	**7**	2.846	2.588	2.548	2.530	2.528	2.506	2.600	2.604
16	**8**	2.870	2.604	2.558	2.534	2.530	2.511	2.622	2.615
16	**9**	2.979	2.692	2.640	2.618	2.604	2.588	2.720	2.699
16	**10**	3.152	2.862	2.794	2.767	2.751	2.741	2.897	2.859
16	**11**	3.420	3.123	3.042	3.014	2.998	2.981	3.168	3.110
16	**12**	3.838	3.510	3.423	3.380	3.366	3.349	3.568	3.487
16	**13**	4.408	4.099	3.995	3.944	3.931	3.916	4.160	4.057
16	**14**	5.319	5.019	4.906	4.844	4.828	4.813	5.082	4.964
16	**15**	6.814	6.608	6.503	6.433	6.413	6.416	6.666	6.561

Tabel 5.5

Response times for ρ = 0.95									
M	**s**	**SQ**	**RM**	**RJ**	**LQES**	**FCFS**	**LQ**	**AGGR**	**ASYMP**
4	**2**	27.971	10.938	10.696	10.744	10.609	10.574	10.937	10.720
4	**3**	∞	9.627	8.733	9.212	8.555	8.466	9.926	8.736
8	**2**	13.475	10.435	10.385	10.356	10.355	10.373	10.415	10.416
8	**3**	12.606	7.443	7.378	7.376	7.357	7.332	7.453	7.424
8	**4**	15.410	6.275	6.069	6.084	6.038	5.997	6.262	6.150
8	**5**	30.468	5.949	5.632	5.639	5.546	5.489	5.999	5.703
8	**6**	∞	6.500	5.955	5.943	5.771	5.697	6.662	5.987
8	**7**	∞	8.837	7.826	7.766	7.434	7.357	9.142	7.734
16	**2**	11.225	10.317	10.342	10.294	10.299	10.306	10.315	10.326
16	**3**	8.412	7.178	7.177	7.170	7.152	7.133	7.180	7.194
16	**4**	7.104	5.695	5.673	5.651	5.657	5.632	5.680	5.696
16	**5**	6.362	4.858	4.804	4.791	4.781	4.757	4.843	4.855
16	**6**	5.981	4.364	4.315	4.250	4.238	4.218	4.351	4.355
16	**7**	5.843	4.078	3.996	3.937	3.931	3.892	4.073	4.061
16	**8**	5.913	3.951	3.848	3.775	3.749	3.738	3.951	3.915
16	**9**	6.179	3.953	3.824	3.751	3.726	3.682	3.963	3.890
16	**10**	6.694	4.092	3.922	3.843	3.803	3.777	4.113	3.986
16	**11**	7.504	4.391	4.159	4.063	4.029	3.995	4.426	4.221
16	**12**	8.846	4.900	4.607	4.488	4.439	4.398	4.965	4.651
16	**13**	11.073	5.758	5.375	5.226	5.168	5.122	5.865	5.397
16	**14**	14.146	7.312	6.785	6.577	6.495	6.459	7.444	6.766
16	**15**	15.704	10.489	9.821	9.497	9.414	9.376	10.640	9.742

Tabel 5.6

From these results we can draw a number of conclusions.

First of all we see from tabel 5.6 that the SQ strategy does not use the capacity efficiently when ρ is large and the number of operators is close to the number of machines. In three cases (indicated by ∞) the system appears to be no longer ergodic. We will come back to this in section 6.

The other 5 strategies are compared in Tabels 5.7 - 5.11. These tables give for $\rho \geq 0.7$ the relative differences between the optimal strategy LQ and the strategies: RM, RJ, LQES and FCFS. For ρ equal to 0.3 and 0.5 the differences are negligible.

Relative percentage difference with LQ policy for ρ = 0.70					
M	**s**	**RM**	**RJ**	**LQES**	**FCFS**
4	**2**	1.733	0.985	1.383	0.852
4	**3**	1.688	0.597	0.935	0.129
8	**2**	1.205	0.870	0.678	0.485
8	**3**	0.838	0.281	0.448	0.337
8	**4**	1.010	0.563	0.451	0.226
8	**5**	0.681	0.368	0.472	0.263
8	**6**	0.461	-0.139	0.231	0.046
8	**7**	0.038	-0.696	-0.231	-0.424
16	**2**	0.298	0.000	0.100	0.000
16	**3**	0.666	0.485	0.365	0.183
16	**4**	0.592	0.330	0.198	0.198
16	**5**	0.271	0.203	0.203	-0.068
16	**6**	0.602	0.269	0.336	-0.067
16	**7**	0.651	0.131	0.196	0.065
16	**8**	0.250	0.187	0.000	0.187
16	**9**	0.295	0.177	-0.059	-0.119
16	**10**	0.277	-0.055	-0.055	-0.111
16	**11**	0.154	0.206	0.154	-0.052
16	**12**	0.236	0.095	-0.095	-0.142
16	**13**	0.516	0.173	0.259	0.086
16	**14**	-0.117	0.155	0.233	-0.117
16	**15**	0.207	0.413	0.207	0.379

Tabel 5.7

Relative percentage difference with LQ policy for ρ = 0.80					
M	**s**	**RM**	**RJ**	**LQES**	**FCFS**
4	**2**	2.382	0.550	0.839	0.550
4	**3**	3.340	1.255	2.510	0.648
8	**2**	1.100	0.656	0.622	0.656
8	**3**	1.900	1.106	1.106	0.386
8	**4**	2.475	1.297	0.898	0.809
8	**5**	2.469	0.851	0.851	0.512
8	**6**	2.331	1.215	0.667	0.445
8	**7**	1.623	0.934	0.586	0.731
16	**2**	0.705	0.177	0.530	0.600
16	**3**	0.912	0.686	0.640	0.458
16	**4**	0.934	0.624	0.521	0.000
16	**5**	0.883	0.333	0.333	-0.056
16	**6**	0.952	0.730	0.226	0.057
16	**7**	1.380	0.612	0.612	0.335
16	**8**	1.332	0.750	0.430	-0.054
16	**9**	1.365	0.864	0.459	0.256
16	**10**	1.271	0.663	0.427	0.238
16	**11**	1.124	0.738	0.392	0.131
16	**12**	0.935	0.392	0.196	0.079
16	**13**	0.724	0.346	0.069	0.104
16	**14**	0.745	0.389	0.060	0.150
16	**15**	0.598	0.150	-0.125	0.225

Tabel 5.8

Relative percentage difference with LQ policy for ρ = 0.90					
M	**s**	**RM**	**RJ**	**LQES**	**FCFS**
4	**2**	2.906	0.959	1.761	0.712
4	**3**	8.800	3.236	5.878	1.586
8	**2**	0.463	0.149	-0.093	0.352
8	**3**	2.419	0.844	1.041	0.175
8	**4**	4.200	1.551	1.496	0.965
8	**5**	5.926	2.075	2.103	1.034
8	**6**	7.327	2.953	2.337	1.030
8	**7**	7.386	3.014	1.936	0.602
16	**2**	-0.169	0.019	-0.377	0.449
16	**3**	0.650	0.235	-0.262	-0.289
16	**4**	1.195	0.789	0.254	0.159
16	**5**	1.977	0.822	0.822	0.430
16	**6**	2.037	0.953	0.688	0.384
16	**7**	3.168	1.648	0.949	0.870
16	**8**	3.571	1.837	0.908	0.751
16	**9**	3.863	1.970	1.146	0.614
16	**10**	4.228	1.897	0.940	0.364
16	**11**	4.547	2.005	1.095	0.567
16	**12**	4.587	2.162	0.917	0.505
16	**13**	4.465	1.977	0.710	0.382
16	**14**	4.104	1.896	0.640	0.311
16	**15**	2.906	1.338	0.264	-0.047

Tabel 5.9

Relative percentage difference with LQ policy for ρ = 0.95					
M	**s**	**RM**	**RJ**	**LQES**	**FCFS**
4	**2**	3.328	1.141	1.582	0.330
4	**3**	12.060	3.057	8.098	1.040
8	**2**	0.594	0.116	-0.164	-0.174
8	**3**	1.491	0.623	0.597	0.340
8	**4**	4.430	1.186	1.430	0.679
8	**5**	7.732	2.539	2.660	1.028
8	**6**	12.354	4.332	4.139	1.282
8	**7**	16.748	5.993	5.267	1.036
16	**2**	0.107	0.348	-0.117	-0.068
16	**3**	0.627	0.613	0.516	0.266
16	**4**	1.106	0.723	0.336	0.442
16	**5**	2.079	0.978	0.710	0.502
16	**6**	3.346	2.248	0.753	0.472
16	**7**	4.561	2.603	1.143	0.992
16	**8**	5.391	2.859	0.980	0.293
16	**9**	6.856	3.713	1.840	1.181
16	**10**	7.698	3.697	1.717	0.684
16	**11**	9.018	3.943	1.674	0.844
16	**12**	10.245	4.537	2.005	0.924
16	**13**	11.046	4.707	1.990	0.890
16	**14**	11.666	4.805	1.794	0.554
16	**15**	10.611	4.531	1.274	0.404

Tabel 5.10

Mean relative percentage difference with LQ policy				
ρ	**RM**	**RJ**	**LQES**	**FCFS**
0.70	0.571	0.254	0.291	0.106
0.80	1.431	0.696	0.584	0.330
0.90	3.662	1.560	1.143	0.541
0.95	6.504	2.695	1.828	0.634

Tabel 5.11

Due to the limited accuracy of the simulations some negative differences are found in spite of the optimality of the LQ policy.

We see that the differences increase with ρ. Also note that the differences between the 5 strategies and the two approximations are always very small if $\rho \leq 0.8$. The differences increase if ρ is increased and are larger if s is close to M if M is not too large.

These results are summarized in tabels 5.12 - 5.14.

Mean and maximal relative differences for $\rho \leq 0.90$					
Percentage difference with AGGR					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	0.64	1.32	1.41	1.64	1.85
Maximal	4.10	10.45	8.87	12.33	14.14
Percentage difference with ASYMP					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	0.79	1.09	1.22	1.41	1.61
Maximal	4.98	2.76	3.32	3.93	4.71

Tabel 5.12

Mean and maximal relative differences for $\rho = 0.95$					
Percentage difference with AGGR					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	0.88	5.10	6.02	7.39	8.09
Maximal	3.45	16.82	17.72	22.98	24.26
Percentage difference with ASYMP					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	3.37	0.86	2.08	2.80	3.46
Maximal	12.48	1.74	5.17	4.81	5.75

Tabel 5.13

Mean and maximal relative differences for all ρ					
Percentage difference with AGGR					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	0.68	1.95	2.18	2.60	2.89
Maximal	4.10	16.82	17.72	22.98	24.26
Percentage difference with ASYMP					
	RM	**RJ**	**LQES**	**FCFS**	**LQ**
Mean	1.22	1.05	1.37	1.64	1.92
Maximal	12.48	2.76	5.17	4.81	5.75

Tabel 5.14

So for $\rho = 0.95$ the Bose-Einstein aggregation gives large errors except for strategy *RM*. For the other 4 strategies the asymptotic approximation is better and even quite good.

6. ERGODICITY

Tabel 5.6 indicated that $\rho < 1$ is not the ergodicity condition for the shortest queue policy as one might have expected. We do not know how to obtain the exact ergodicity condition. The following reasoning gives some insight in the problem.

Let us consider the case $s = M-1$. And let us assume a preemptive resume policy is used with priority to the machine with the lower number, so machine 1 has priority over all other machines, machine 2 has priority over 3 upto M, etc. Since there are $M-1$ operators there is always an operator available for any of the machines 1 upto $M-1$. So these machines all behave as $M \mid M \mid 1$ queues with arrival rate and utilization $\sigma := s\rho/M = (M-1)\rho/M$.

The fraction of time an operator is available for machine M is just the fraction of time that at least one of the first $M-1$ machine queues is empty, which is equal to $1-\sigma^{M-1}$. The arrival rate to machine M equals σ as well. So the system will be ergodic if

$$\sigma < 1 - \sigma^{M-1} . \tag{5.1}$$

The ergodicity conditions for M = 4,8 and 16 are given in tabel 6.1.

Ergodicity condition preemptive priorities			
	M=4	M=8	M=16
ρ_{max}	0.910	0.910	0.930

Table 6.1

If preemptions are not allowed the behaviour of the system is more complicated. But it seems clear that in that case as well $\rho < 1$ will not be sufficient either.
In the case of the shortest queue policy there are no preemptions and no priorities. But once one of the machine queues has become long the operators are going to give priority to the other (shorter!) machine queues. So with one long queue the system behaves as the nonpreemptive version of the system described above. Intuitively this explains to us why $\rho < 1$ is not sufficient.

Our conjecture is that $\rho < 1$ is sufficient for the policies LQ, FCFS, LQES and RJ and not for RM and SQ.

7. NON-EXPONENTIAL SERVICE TIMES

The asymptotic approximation given in section 4 suggests a number of straightforward generalizations for non-exponential service times. Note that in (4) we can replace the waiting times for the $M\,|M\,|\,1$ and $M\,|M\,|\,s$ queues by the waiting time in the $M\,|\,G\,|\,1$ and one of the approximations for the waiting time in the $M\,|\,G\,|\,s$ queues, see e.g. Tijms [2]. Then, as one easily verifies, the same α can be used as in (6) for the exponential service times.

Tables 7.1 and 7.2 give the simulation results for Erlang-2 and Erlang-4 distributed service times. We compare the results for the strategies RM, RJ and LQ with the asymptotic approach. For the $M\,|\,G\,|\,s$ queue we used the approximation given in equations (4.196) and (4.200) in Tijms[2]:

$$W_{M\,|G\,|s} = \left[(1-\rho)\gamma s + \rho(1+c^2)/2 \right] W_{M\,|M\,|s}\,,$$

with

$$\gamma = (1-c^2)\frac{1}{s+1} + c^2\frac{1}{s}\,,$$

and c^2 the squared coefficient of variation (1/2 for Erlang-2 and 1/4 for Erlang-4).

Response times for Erlang-2 service times						
M	S	ρ	**RM**	**RJ**	**LQ**	**ASYMP**
8	**6**	0.700	1.882	1.881	1.871	1.908
8	**6**	0.950	5.180	4.773	4.564	4.763
8	**7**	0.700	2.203	2.201	2.198	2.223
8	**7**	0.950	6.951	6.200	5.830	6.068
16	**13**	0.700	1.990	1.995	1.994	2.004
16	**13**	0.950	4.604	4.315	4.153	4.308
16	**14**	0.700	2.190	2.187	2.189	2.195
16	**14**	0.950	5.758	5.377	5.170	5.334
16	**15**	0.700	2.427	2.431	2.430	2.436
16	**15**	0.950	8.131	7.657	7.399	7.563

Tabel 7.1

Response times for Erlang-4 service times						
M	S	ρ	**RM**	**RJ**	**LQ**	**ASYMP**
8	**6**	0.700	1.743	1.735	1.731	1.760
8	**6**	0.950	4.491	4.147	4.006	4.151
8	**7**	0.700	2.004	2.000	2.003	2.021
8	**7**	0.950	5.933	5.334	5.059	5.234
16	**13**	0.700	1.829	1.828	1.829	1.838
16	**13**	0.950	3.995	3.771	3.643	3.764
16	**14**	0.700	1.984	1.989	1.990	1.996
16	**14**	0.950	4.964	4.652	4.488	4.617
16	**15**	0.700	2.198	2.191	2.195	2.197
16	**15**	0.950	6.927	6.534	6.354	6.474

Tabel 7.2

As we see the result are just as good as in the exponential case. The Asymptotic approximation is quite close to the results for the Random Job policy.

A more primitive approach is to simply multiply the waiting times for the exponential case by the expected residual service time divided by the mean service time. If we do this with the simulation results for the exponential service time then the results are between 0 and 2 percent below the simulation results for the Erlang service times.

It might be worthwhile to find good approximations for the differences between the various strategies and the differences with the approximations.

References

[1] DURLINGER, P.P.J. and J.H.J. GEURTS, Determining work order throughput times in dual constrained production departments, Eindhoven University of Technology, Department of Industrial Engineering, Working paper BDK/KBS/89-13, December 1989.

[2] TIJMS, H.C., Stochastic Modelling and Analysis: A Computational Approach. Wiley, Chichester, 1986, edition 1988.

[3] WAL, J. VAN DER, The longest queue policy is optimal for the assignment of identical operators to different machines, Eindhoven University of Technology, Department of Mathematics and Computing Science, Memorandum COSOR, 1994.

List of COSOR-memoranda - 1993

Number	Month	Author	Title
93-01	January	P. v.d. Laan C. v. Eeden	Subset selection for the best of two populations: Tables of the expected subset size
93-02	January	R.J.G. Wilms J.G.F. Thiemann	Characterizations of shift-invariant distributions based on summation modulo one.
93-03	February	Jan Beirlant John H.J. Einmahl	Asymptotic confidence intervals for the length of the shortt under random censoring.
93-04	February	E. Balas J. K. Lenstra A. Vazacopoulos	One machine scheduling with delayed precedence constraints
93-05	March	A.A. Stoorvogel J.H.A. Ludlage	The discrete time minimum entropy H_∞ control problem
93-06	March	H.J.C. Huijberts C.H. Moog	Controlled invariance of nonlinear systems: nonexact forms speak louder than exact forms
93-07	March	Marinus Veldhorst	A linear time algorithm to schedule trees with communication delays optimally on two machines
93-08	March	Stan van Hoesel Antoon Kolen	A class of strong valid inequalities for the discrete lot-sizing and scheduling problem
93-09	March	F.P.A. Coolen	Bayesian decision theory with imprecise prior probabilities applied to replacement problems
93-10	March	A.W.J. Kolen A.H.G. Rinnooy Kan C.P.M. van Hoesel A.P.M. Wagelmans	Sensitivity analysis of list scheduling heuristics
93-11	March	A.A. Stoorvogel J.H.A. Ludlage	Squaring-down and the problems of almost-zeros for continuous-time systems
93-12	April	Paul van der Laan	The efficiency of subset selection of an ε-best uniform population relative to selection of the best one
93-13	April	R.J.G. Wilms	On the limiting distribution of fractional parts of extreme order statistics

Number	Month	Author	Title
93-14	May	L.C.G.J.M. Habets	On the Genericity of Stabilizability for Time-Day Systems
93-15	June	P. van der Laan C. van Eeden	Subset selection with a generalized selection goal based on a loss function
93-16	June	A.A. Stoorvogel A. Saberi B.M. Chen	The Discrete-time H_∞ Control Problem with Strictly Proper Measurement Feedback
93-17	June	J. Beirlant J.H.J. Einmahl	Maximal type test statistics based on conditional processes
93-18	July	F.P.A. Coolen	Decision making with imprecise probabilities
93-19	July	J.A. Hoogeveen J.K. Lenstra B. Veltman	Three, four, five, six or the Complexity of Scheduling with Communication Delays
93-20	July	J.A. Hoogeveen J.K. Lenstra B. Veltman	Preemptive scheduling in a two-stage multiprocessor flow shop is NP-hard
93-21	July	P. van der Laan C. van Eeden	Some generalized subset selection procedures
93-22	July	R.J.G. Wilms	Infinite divisible and stable distributions modulo 1
93-23	July	J.H.J. Einmahl F.H. Ruymgaart	Tail processes under heavy random censorship with applications
93-24	August	F.W. Steutel	Probabilistic methods in problems of applied analysis
93-25	August	A.A. Stoorvogel	Stabilizing solutions of the H_∞ algebraic Riccati equation
93-26	August	R. Perelaer J.K. Lenstra M. Savelsbergh F. Soumis	The Bus Driver Scheduling Problem of the Amsterdam Transit Company
93-27	August	J.M. van den Akker C.P.M. van Hoesel M.W.P. Savelsbergh	Facet Inducing Inequalities for Single-Machine Scheduling Problems
93-28	September	H.J.C. Huijberts H. Nijmeijer	Dynamic disturbance decoupling of nonlinear systems and linearization

Number	Month	Author	Title
93-29	September	L.C.G.J.M. Habets	Testing Reachability and Stabilizability of Systems over Polynomial Rings using Gröbner Bases
93-30	September	Z. Liu J.A.C. Resing	Duality and Equivalencies in Closed Tandem Queueing Networks
93-31	November	A.G. de Kok J. van der Wal	Assigning identical operators to different machines