Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

APPROXIMATE SOLUTIONS FOR M/G/1 FORK/JOIN SYNCHRONIZATION

Alexander Thomasian
Asser N. Tantawi

IBM Research Division
Thomas J. Watson Research Center
Yorktown Heights, NY 10598

ABSTRACT

Approximation techniques are developed to evaluate the performance of symmetric fork-join synchronization delays for K $M/G/1$ queues. For a server utilization ρ, the mean response time for fork-join requests is expressed as the sum of the mean response time at one of the queues and the mean synchronization delay as follows: $R_K^{F/J}(\rho) = R_1(\rho) + F_K \alpha_K(\rho)\sigma_1(\rho)$, where F_K is obtained from the previous equation at $\rho = 0$ (since $\alpha_K(0) \stackrel{\wedge}{=} 1$). $R_1(\rho)$ and $\sigma_1(\rho)$ are the mean and the standard deviation of response time at any one of the queues, respectively, and $\alpha_K(\rho)$ is a low-degree service-time distribution dependent polynomial in ρ, whose coefficients are determined from simulation results. We also use simulation results to show that when fork-join requests share the servers with local requests, a good approximation (and an upper bound) to the fork-join response time is obtained by treating the components of fork-join response time as independent, i.e., the mean fork-join response time can be approximated by the expected value of the maximum of the response times at the K queues.

1 INTRODUCTION

We consider a fork-join queueing system where an arrival generates K requests for the K servers of the system. The fork-join response time is then determined by the completion of all requests generated by the arrival. Note that this response time denoted by $R_K^{F/J}(\rho)$ at server utilization ρ is different from the maximum of K independent response times (denoted by $R_K^{max}(\rho)$), which is in fact an upper bound to the mean fork-join response time as shown by Nelson and Tantawi (1988).

The present study was motivated by the need to analyze the performance of disk arrays operating in degraded mode by Thomasian and Menon (1994), where an access to the failed disk requires accesses to corresponding blocks on all surviving disks to recreate the data block being accessed. Next we discuss the two papers on fork-join synchronization, which are most directly related to this study.

A system with Poisson arrivals and exponential service times is considered by Nelson and Tantawi (1988). A scaling approximation is introduced based on the observation that both the lower and upper bounds of $R_K^{F/J}(\rho)$ grow at the same rate as a function of K. This observation leads to the expression $R_K^{F/J}(\rho) = S_K(\rho)R_2(\rho)$, where $S_K(\rho) = \alpha(\rho) + (1 - \alpha(\rho))H_K/H_2$, with $H_K = \sum_{k=1}^{K} 1/k$ and $R_2(\rho) = (1.5 - \rho/8)R_1(\rho)$. The coefficients of the polynomial $\alpha(\rho)$ are estimated using simulation results as $\alpha(\rho) \simeq 4\rho/11$, leading for $K \geq 2$ to

$$R_K^{F/J}(\rho) = [\frac{H_K}{H_2} + \frac{4}{11}(1 - \frac{H_K}{H_2})\rho]R_2(\rho). \quad (1.1)$$

More recently, an interpolation approximation between light and heavy traffic is introduced by Varma and Makowski (1994) for fork-join queueing systems with general inter-arrival and service times. The light traffic approximation is based on computing the mean fork-join response time and its derivative, the latter of which is especially involved. The heavy traffic approximation is based on solving an instance of the problem for $K = 2$ and an agreement between the light traffic derivative and the heavy traffic limit. While the proposed heuristic approximations are shown to be acceptably accurate (through validation against simulation results), the approximate solution method proposed in this paper is also of interest due to the following reasons:

1. It is based on fewer approximations and requires less sophisticated analytic techniques, which makes the methodology accessible to a wider range of performance modeling practitioners.

2. It can be applied to any service time distribution, while analytic expressions for service distribution time specific parameters are required by the other methed, which is the reason that it has been applied only to the exponential, Erlang-2, and hyperexponential distributions.
3. Our method works best for larger values of K, while the error introduced by the other method tends to increase with K (a maximum value of $K = 15$ is considered in this study).
4. The method described in this paper is similar in nature to Neslon and Tantawi (1988) and requires initial simulations, while the other method does not.

We consider a fork-join queueing system with K servers, Poisson arrivals (with parameter λ) and general service times with PDF $B(t)$, $t > 0$. The ith moment (resp. variance) of service time is denoted by b_i (resp. σ_B^2). The moments of response time at individual queues are that of an $M/G/1$ queueing system with the mean (resp. standard deviation) denoted by $R_1(\rho)$ (resp. $\sigma_1(\rho)$), where $\rho = \lambda b_1$ is the utilization of each server. The coefficient of variation of response time is $c_R \triangleq \sigma_1(\rho)/R_1(\rho)$.

In Section 2 we propose a methodology which uses the expression $R_K^{F/J}(\rho) = R_1(\rho) + F_K \alpha_K(\rho)\sigma_1(\rho)$ as a starting point, with $F_K = (b_K^{max} - b_1)/\sigma_B$, where b_K^{max} is the expected value of the maximum of K service times given as $b_K^{max} = \int_0^\infty [1 - B^K(t)]dt$ (note that $B^K(t)$ is the distribution for the maximum of K service times). Given that $R_1(\rho)$ and $\sigma_1(\rho)$ can be derived analytically, we use simulation to obtain an adequate number of data points for $R_K^{F/J}(\rho)$, which are then used to obtain $\alpha_K(\rho)$ by using surface-fitting. Given $\alpha_K(\rho)$ the above equation can be used to compute $R_K^{F/J}(\rho)$ at a very low cost.

Next in Section 3 we consider fork-join queueing systems with balanced interfering traffic, i.e., local requests with equal arrival rates at the servers. Conclusions appear in Section 4.

2 AN APPROXIMATION FOR FORK-JOIN RESPONSE TIMES

Before considering general service times, we first consider the Markovian queueing system by Nelson and Tantawi (1988) with Poisson arrivals and exponential service times. Eq. (1.1) in Section 1 can be rewritten as:

$$R_K^{F/J}(\rho) = R_1(\rho) + \left[H_K - 1 - \frac{59(H_K - 1) - 13}{132}\rho + \frac{H_K - 1.5}{33}\rho^2 \right]\sigma_1(\rho) \qquad (2.1)$$

Note that $\sigma_1(\rho) = R_1(\rho)$ since the individual queues are M/M/1 and that H_K denotes the expansion factor for the response time due to synchronization delays at $\rho = 0$. The synchronization delay in this case normalized by $(H_K - 1)\sigma_1(\rho)$ is

$$\alpha_K(\rho) = \frac{R_K^{F/J}(\rho) - R_1(\rho)}{(H_K - 1)\sigma_1(\rho)} \qquad (2.2)$$

$$= 1 - (59 - \frac{13}{H_K - 1})\frac{\rho}{132} + (1 - \frac{0.5}{H_K - 1})\frac{\rho^2}{33}$$

In Figure 2.1 we plot $\alpha_K(\rho)$ versus ρ for $K = 2, 5, 10$, and 15 (the four cases considered by Varma and Makowski (1994)) and $K = 32, 48$, and 64 (new simulation results showed that Eq. (1.1) estimates the mean fork-join response time for $K > 32$ very accurately as well). In fact, the "lettered dots" connecting the graphs are the midpoints of confidence intervals with a width of 1% at 95% confidence level (slightly wider at the highest utilization for $K \geq 32$ to limit the simulation time). Note that $\alpha_K(\rho)$ decreases with increasing ρ, which means that the normalized synchronization delay decreases with increasing ρ (this is not generally true). In addition the graphs are linear in K becoming less distinguishable as K increases until a limiting slope is reached. This is also true for the other distributions considered in this study. This is attributable to the fact that the increase in the synchronization delay is just reflected by F_K and that $\alpha_K(\rho)$ has a negligible additional effect for small increments in K when K is already large.

We next propose the following expression for $R_K^{F/J}(\rho)$ for a general service time distribution

$$R_K^{F/J}(\rho) = R_1(\rho) + F_K \sigma_1(\rho)\alpha_K(\rho) \qquad (2.3)$$

This equation is motivated by several approximations and upper bounds of similar form discussed in Chapter 4 by David (1981). Given that W denotes the mean waiting time in each queue, then $R_1(\rho) = b_1 + W$ with $W = \lambda b_2/(2(1 - \rho))$ and $\sigma_R^2(\rho) = \sigma_B^2 + \sigma_W^2 + 2Wb_1$ with $\sigma_W^2 = W^2 + \lambda b_3/(3(1 - \rho))$ (see Takagi (1991)). F_K captures the increase in response time at $\rho = 0$ and is consequently given as

$$F_K = \frac{b_K^{max} - b_1}{\sigma_B}. \qquad (2.4)$$

Note that $F_K = H_K - 1$ in the case of the exponential distribution and $F_K = \gamma\sqrt{6}\,(\ln(K) - 1)/\pi$ in the case

of the extreme value distribution (see e.g., Johnson and Kotz (1970)). where $\gamma = 0.57722$ is Euler's constant. Given F_K we have

$$\alpha_K(\rho) = \frac{R_K^{F/J}(\rho) - R_1(\rho)}{F_K \sigma_1(\rho)} \tag{2.5}$$

In the case of the exponential service times, the distribution of response time at individual servers is also exponential $\sigma_1(\rho) = R_1(\rho)$, such that $R_K^{max}(\rho) = H_K R_1(\rho)$. Since it is shown by Neslon and Tantawi (1999) that

$$R_K^{F/J}(\rho) \le R_K^{max}(\rho) = H_K R_1(\rho),$$

it follows that $\alpha_K(\rho) \le 1$. For $K = 2$ we simply have $\alpha_2(\rho) = 1 - \rho/4$. If we ignore the non-linear term in Eq. (2.2) (or equivalently postulate that $\alpha_K(\rho)$ is linear for $K > 2$) then it is observed that the absolute value of the slope of the line increases with increasing K. In fact this behavior was observed for all other distributions considered in this study. It can be observed from Figures 2.1-2.4 that $\alpha_K(\rho)$ attains a value smaller than one as $\rho \to 1$ and $K \to \infty$, which is observed to get smaller as c_B decreases. For fixed service times $F_K = 1$ and $\alpha_K(\rho) = 0$ for $\rho > 0$.

In the case of the exponential distribution and other distributions with a small coefficient of variation it can be observed from Figures 2.1, 2.2, and 2.4 that $\alpha_K(\rho)$ decreases with increasing ρ. This behavior is not consistent, however, and is violated in the case of hyperexponential distributions with high variability (see Figure 2.3).

We next test the suitability of Eq. (2.3) in estimating $R_K^{F/J}(\rho)$ for the various distributions considered by Varma and Makowski (1994) and one additional distribution ($b_1 = 1$ in all cases). B_K^{max} required for computing F_K can be computed easily from $B_K^{max} = \int_0^\infty [1 - B^K(t)]dt$ for some distributions such as exponential, Erlang, and hyperexponential distributions. Numerical integration techniques are required in other cases, while simulation can be used alternatively to estimate $R_K^{F/J}$ with $\lambda \to 0$. The values of $R_K^{F/J}(\rho)$ used in this study are based on simulation results given by Varma and Makowski (1994) and provided by our own simulator.

Figure 2.2 shows $\alpha_K(\rho)$ versus ρ for the Erlang-2 distribution with $\sigma_B = 1/\sqrt{2}$. b_K^{max} is given by Varma and Makowski (1994)

$$b_K^{max} = \frac{1}{\mu} \sum_{n=1}^{K} \binom{K}{n} (-1)^{n-1} \sum_{m=1}^{n} \binom{n}{m} \frac{m!}{2n^{m+1}} \tag{2.6}$$

Hence $F_K = \sqrt{2}\,(b_K^{max} - 1)$. Trends previously observed for the exponential distribution also hold in this case, e.g., $\alpha_K(\rho)$ decreases with ρ and K but the graphs become indistinguishable as K increases.

The graphs for the hyperexponential distribution $B(t) = 1 - (p_1 e^{-\mu_1 t} - p_2 e^{-\mu_2 t}), t \ge 0$, with $0 < p_1 < 1$, $p_2 = 1 - p_1$ and $\mu_1 \ne \mu_2$, considered by Varma and Makowski (1994) appear in Figure 2.3. The computations are simplified for $p_1/\mu_1 = p_2/\mu_2 = 0.5$, which implies $\mu_1 + \mu_2 = 2$ for $b_1 = 1$. The following parameter settings are used with $\mu_1 = 0.1$, $\mu_2 = 1.9$, $p_1 = 0.05$, and $p_2 = 0.95$. The variance of service time is $\sigma_B^2 = 2/(\mu_1\mu_2) - 1 = 41.10$ and $c_B = 6.4$. b_K^{max} is given by Varma and Makowski (1994):

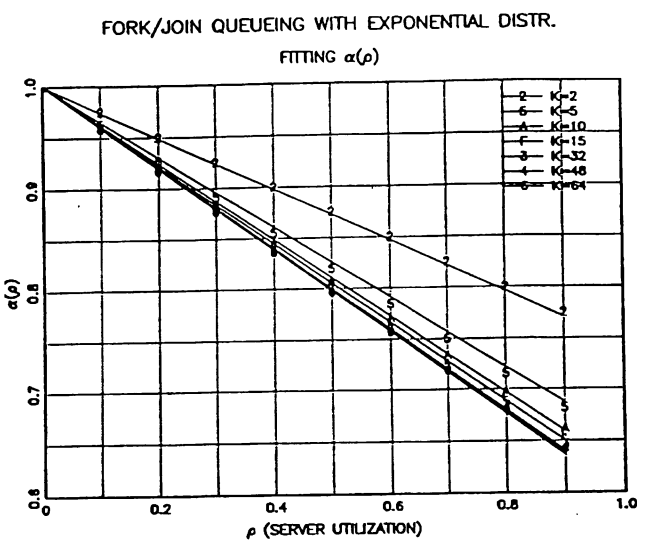
$$b_K^{max} = \sum_{n=1}^{K} (-1)^{n+1} \sum_{m=0}^{n} \binom{n}{m} \frac{p_1^m p_2^{n-m}}{m\mu_1 + (n-m)\mu_2} \tag{2.7}$$

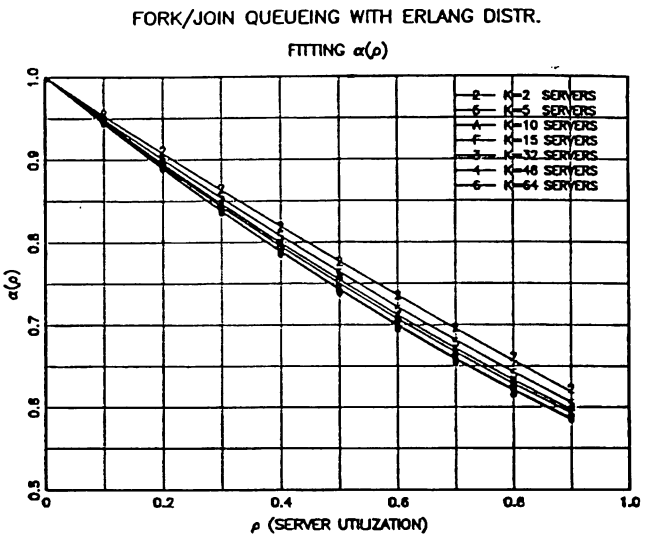
from which F_K follows. In contrast to previous cases, $\alpha_K(\rho)$ increases initially with ρ and this increase is the highest for $K = 2$, i.e., the variability in service times has the most effect for smaller values of K. After achieving a maximum $\alpha_K(\rho)$ decreases with ρ and for increasing values of K the maximum is attained at smaller values of ρ, until the graphs become almost linear.

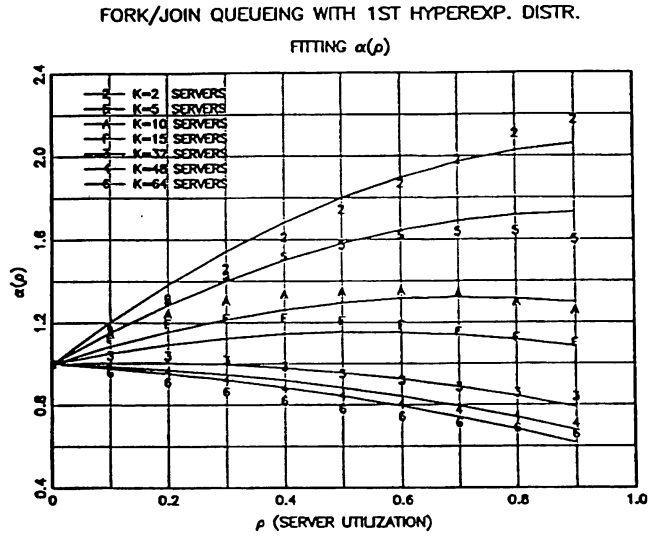
We also consider a second hyperexponential distribution with less variability with the following parameters $\mu_1 = 0.5$, $\mu_2 = 1.5$, $p_1 = 0.25$, and $p_2 = 0.75$ yielding $b_1 = 1$, $\sigma_B^2 = 1.67$, and $c_B = 1.3$. It can be observed from Figure 2.4 that the corresponding values of $\alpha_K(\rho)$ in this case are much smaller than those for the previous hyperexponential distribution and that they attain a linear form in ρ for much smaller values of K than in Figure 2.3.

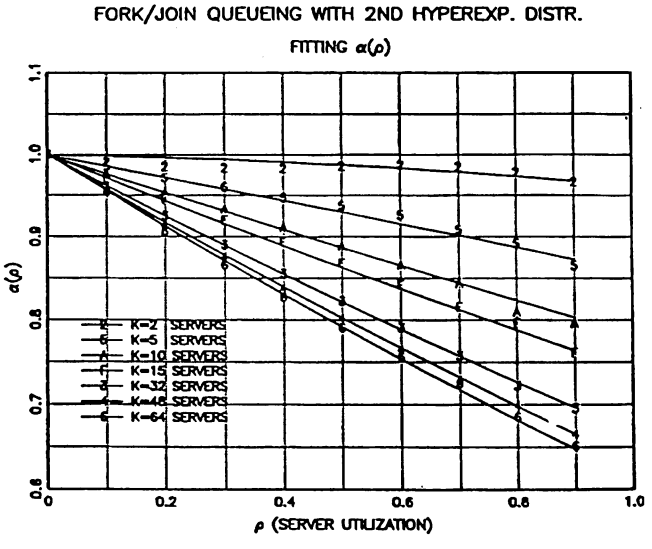
We carried out the following experiments to obtain $\alpha_K(\rho)$ for the forementioned distributions. We first plotted $\alpha_K(\rho)$ (given by Eq. (2.5)) for a set of values for K versus $0 < \rho \le 0.9$. We then used surface-fitting to derive the function $\alpha_K(\rho)$.

In the case of the exponential service time distribution, referring back to Figure 2.1, $\alpha_K(\rho)$ can obviously be expressed quite accurately as a linear function. The surface-fitting operation of the AGSS package developed at IBM Research (see e.g., Lane and Welch (1987)) yielded several functions of which $\alpha_K(\rho) = 1 - (a - b/K)\rho$ with $a = 0.409$ and $b = 0.310$ turned out to be the most simple satisfactory fit. Note that for larger values of K $\alpha_K(\rho) \simeq 1 - a\rho$ and since F_K increases very slowly with K, the values of $R_K^{max}(\rho)$ become indistinguishable as K is gradually increased. It follows that $\lim_{K\to\infty} R_K^{max}(\rho) \simeq R_1(\rho) + (\ln(K) + \gamma - 1)(1 - a\rho)\sigma_1(\rho)$ with $a = 0.409$.

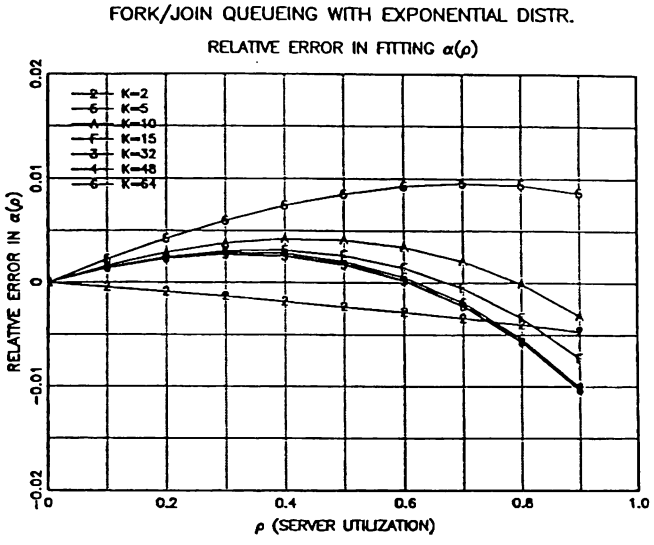


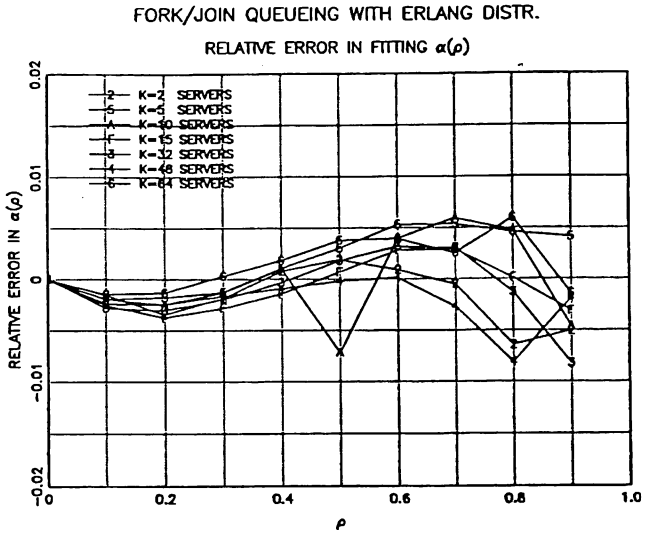


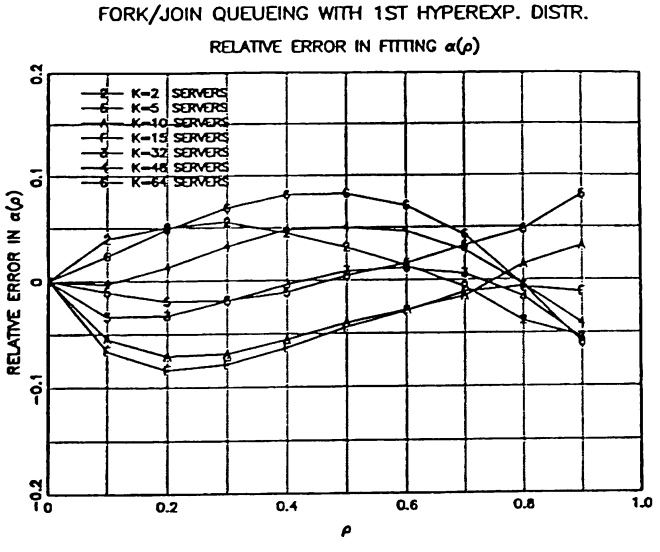




Figures 2.1-4. The function $\alpha_K(\rho)$ versus server utilization (ρ) for varying number of servers (K) for the four distributions. The "lettered dots" correspond to midpoint of a 95% confidence interval with a 1% width obtained by simulation and the graphs are fitted with equations appearing in the text.







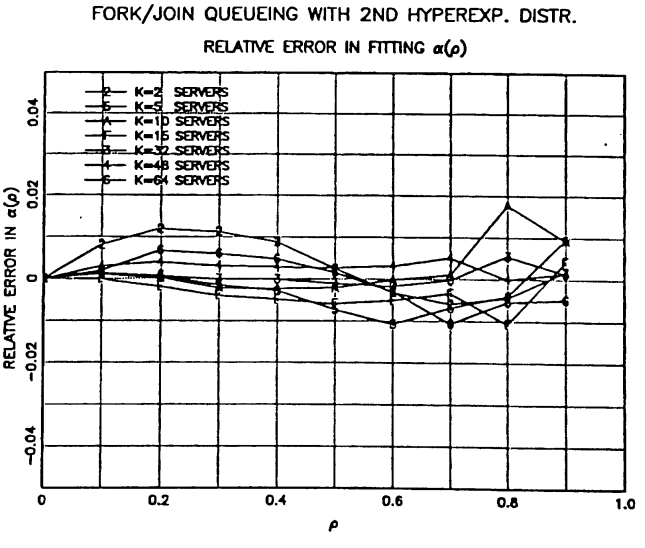


Figure 2.5-8. Relative error in the fitted values of $\alpha_K(\rho)$ with respect to the original values estimated by simulation versus server utilization (ρ) for varying number of servers (K) for the four distributions.

In the case of the two hyperexponential distributions a limited degree of experimentation with surface-fitting yielded

$$\alpha_K(\rho) = 1 + a\rho + b\rho^2 - \rho(c + d\rho)\frac{\log_2 K}{K + e} \quad (2.8)$$

Note that $\lim_{K\to\infty} \alpha_K(\rho) = 1 + a\rho + b\rho^2$. The coefficients for the first and second hyperexponential distributions are respectively:

$a = -0.5783$, $b = -0.1395$, $c = -4.3110$, $d = 1.4629$, and $e = 0.4068$;

$a = 0.08879$, $b = -0.4902$, $c = 59.0712$, $d = -12.3882$, and $e = 596.4158$.

It turns out the above function is also a good fit for the Erlang distribution in which case surface-fitting yielded: $a = -0.4404$, $b = 0.0317$, $c = 5.6817$, $d = -4.1722$, and $e = 172.7550$.

Figures 2.1-2.4 are plots of the function $\alpha_K(\rho)$ versus ρ for different values of K for the four distributions (the lettered dots are simulation results). In Figures 2.5-2.8 we have plotted the relative error (with respect to the value obtained by simulation) in estimating $\alpha_K(\rho)$ through the surface-fitting process. This is tantamount to a smaller error in $R_K^{F/J}(\rho)$, especially when the coefficient of variation of response time (c_R) is small. It is observed from the graphs that the relative error is in the range of few percentage points in all cases, with the first hyperexponential distribution being an exception. This is attributable to the different forms of the $\alpha_K(\rho)$ graphs for different values of K for this particular distribution. In fact a better fit could be obtained by considering two separate regions for smaller and higher values of K. Finally, it can be shown by plotting the mean fork-join response time ($R_K^{F/J}(\rho)$ versus ρ, that by applying this "inverse transformation" we are able to estimate the mean fork-join response time quite accurately.

It is observed from Figures 2.2-2.4 that $\alpha_K(\rho)$ for $K \to \infty$ attains a linear form and also demonstrates a limiting behavior. There are two benefits: (i) linear interpolation is the least difficult; (ii) the limiting behavior makes it possible to *extrapolate* performance for even higher values of K.

Based on the above experiments we next outline a simple methodology to obtain the required parameters for Eq. (2.3).

1. For a given service time distribution compute $R_1(\rho)$ and $\sigma_1(\rho)$ as ρ is varied.
2. Compute F_K using Eq. (2.4) after obtaining b_K^{max} for an appropriate set of values for K.
3. Use simulation to estimate $R_K^{max}(\rho)$ for the set of values for K and ρ.
4. Obtain an adequate number of data points to compute the coefficients in Eq. (2.8), by solving the associated non-linear equations.

This technique is expected to be particularly robust, yielding very accurate estimates of $R_K^{F/J}$ when the coefficient of variation c_R is not very large and the interest is in higher values of K. While Eq. (2.8) was found to be applicable to the several distributions considered in this study (the exponential distribution is a special case), other forms might be more suitable for other distributions. Thus the major contribution of this work is the methodology, rather than the specific equations derived in the process.

3 PERFORMANCE ANALYSIS OF SYSTEM WITH INTERFERING TRAFFIC

A symmetric fork-join system with ordinary requests uniformly distributed over the servers in addition to fork-join requests is considered. To study the behavior of the system the arrival rate of requests as well as the intensity of local requests is varied, such that the contribution of ordinary requests to server utilization is in the range 0.1 to 0.9. The service time distribution of local requests is assumed to be the same as fork-join requests in this discussion.

Figures 3.1-3.4 shows the mean response time of fork-join (not overall) requests versus server utilization for different intensities of local traffic (specified as a fraction of server utilization) in a system with $K = 8$. The four figures correspond to the exponential distribution, the Erlang and the two hyperexponential distributions in Section 2, respectively. The graph in Figure 3.1 which is denoted by L is the approximate solution by Nelson and Tantawi (1988) for the case of no local traffic (this approximation is very accurate and simulation results are therefore not plotted in this case). We also plot the maximum of the K response times (denoted by H) which in the exponential case is simply $R_K^{max}(\rho) = H_K R_1(\rho)$. It can be observed from Figure 3.1 for the exponential case that the approximation by Nelson and Tantawi (1988) becomes less accurate for higher levels of interfering traffic and in fact for such levels of interfering traffic the maximum obtained using the independence assumption is a good approximation.

The fact that the response times at different queues tend to be independent from each other is due to the fact that the distribution of queue lengths encountered by fork-join requests (which determine

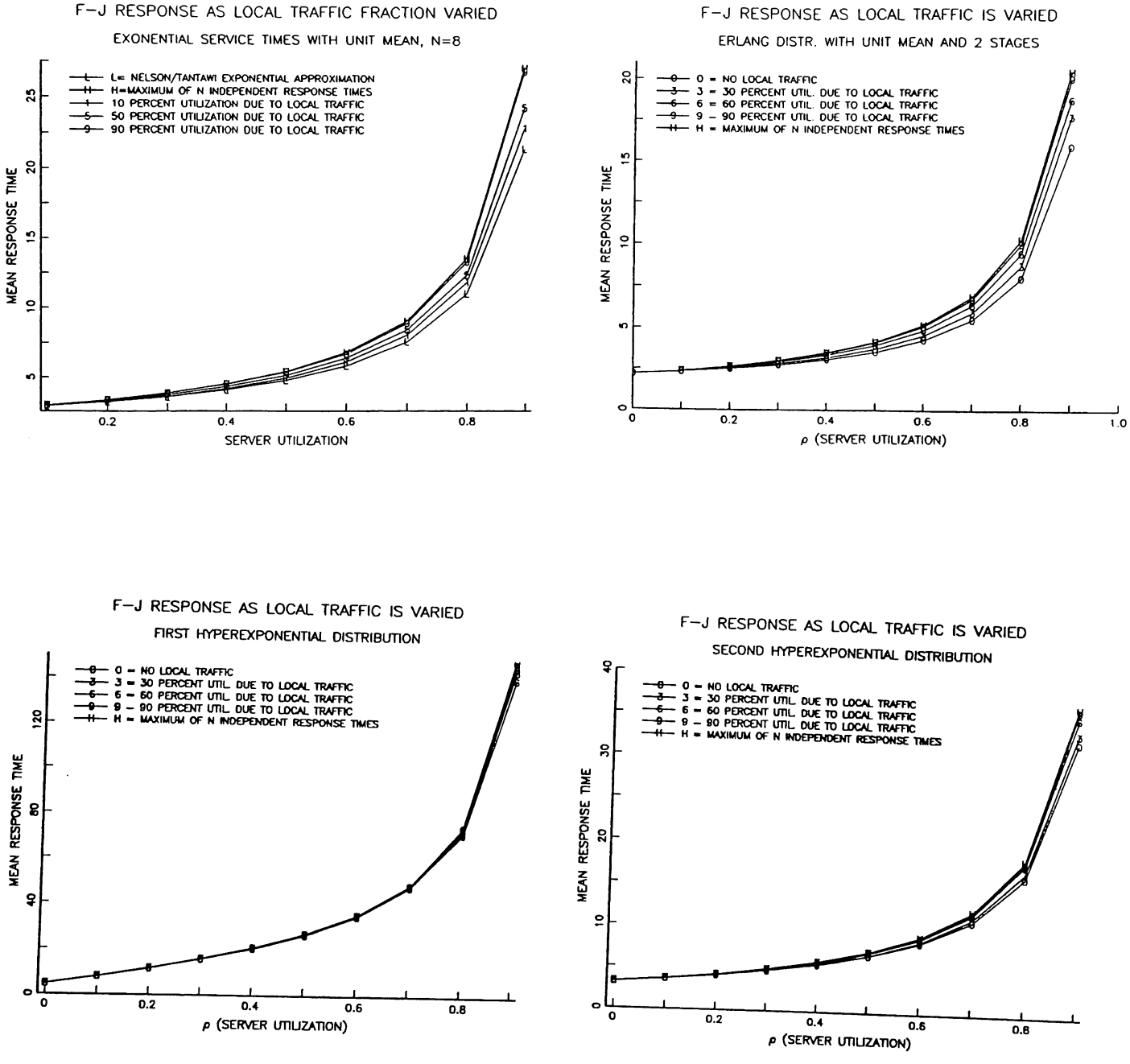


Figure 3.1-4. The effect of varying the rate of local requests on mean fork-join response time versus server utilization (ρ) for $K = 8$. servers.

the individual fork-join response times) tend to be independent of each other when the fraction of such requests processed by each server is small with respect to local requests.

A similar conclusion can be drawn from Figures 3.2-3.4. To obtain the upper bound to fork-join response time in thes cases, we first derive the response time Laplace transform at each $M/G/1$ queue, invert it, and then evaluate $R_K^{max} = \int_0^\infty [1 - R^K(t)]dt$, which can be done easily since it consists of exponentials.

4 CONCLUSIONS

We have presented a methodology to estimate the mean response time for fork-join requests ($R_K^{F/J}(\rho)$). This response time is expressed as the sum of the mean and the standard deviation multiplied by a distribution specific coefficient, which also depends on server utilization. The latter is the product of a coefficient F_K related to the maximum of K service times and a polynomial $\alpha_K(\rho)$. While $\alpha_K(\rho)$ may have a rather irregular shape for smaller values of K and higher values of the coefficient of variation of service time, it tends to a linear function in ρ with a negative slope. In fact a limiting behavior at K_{limit} is observed as K increases. The above observations have two practical implications: (i) interpolation can be used to estimate $R_K^{F/J}(\rho)$ after appropriate curve-fitting; (ii) $R_K^{F/J}(\rho)$ can be extrapolated beyond K_{limit}.

In the case of a fork-join system with interfering traffic the intensity of this traffic affects the fork-join response time. When fork-join requests constitute a negligible fraction of server utilization, then the components of fork-join response time can be treated as independent requests. The independence assumption yields an upper bound to response time which is also a good approximation. Note that this solution method is also applicable when the local traffic is unbalanced.

REFERENCES

David, H. A. 1981. *Order Statistics, 2nd ed.* John-Wiley and Sons.

Johnson, N. L. and Kotz, S. 1970. *Distributions in Statistics: Continuous Univariate Distributions-1,* John-Wiley and Sons, New York, NY.

Lane, T. P. and Welch, P. D. 1987. "The Integration of a Menu-Oriented Graphical Statistical System with its Underlying General Purpose Language" *Computer Science and Statistics: Proc. 19th Symp. on the Interface,* Philadelphia, PA, pp. 267-273.

Nelson, R. and Tantawi, A. N. 1988. "Approximate analysis of fork/join synchronization in parallel queues," *IEEE Trans. on Computers 37,*6, 739-743.

Takagi, H. 1991. *Queueing Analysis, Vol. 1: Vacation and Priority Systems, Part 1,* North-Holland.

Thomasian, A. and Menon, J. 1994. "Performance analysis of RAID5 disk arrays with a vacationing server model for rebuild mode operation," *Proc. 10th IEEE Int'l Conf. on Data Engineering,* Houston TX, pp. 111-119.

Varma, S. and Makowski, A. M. 1994. "Interpolation approximations for symmetric fork-join queues," *Perform. Eval. 20,*1-3, 145-165.

AUTHOR'S BIOGRAPHIES

ALEXANDER THOMASIAN received the Ph.D. degree in computer science from the University of California at Los Angeles. He has been a member of research staff in the Systems Analysis Department at the IBM T. J. Watson Research Center since 1985. He was a faculty member at Case Western University and the University of Southern California and a senior staff scientist at Burroughs Corp. He is interested in the performance analysis and design of parallel and distributed systems and disk array performance. Dr. Thomasian has served on the program committees of the ACM Sigmetrics, IEEE Data Engineering and Distributed Computing Conferences. He is a senior member of IEEE and a member of ACM.

ASSER N. TANTAWI received the Ph.D. degree in computer science from Rutgers University in 1982. He joined the IBM Thomas J. Watson Research Center in 1982, where he is currently manager of Systems Connectivity Performance in the Communications Systems Department. His fields of interest include performance modeling, queuing theory, load balancing, parallel processing, reliability modeling, high-speed networking, and Intelligent Vehicle Highway Systems. Dr. Tantawi is a senior member of IEEE and a member of ACM and ORSA/TIMS, and served as an ACM National Lecturer (1984-1988), and as a guest editor for Performance Evaluation.