An Approximate Model for Field Service Territory Planning

ARTHUR V. HILL

SALVATORE T. MARCH

CHRISTOPHER J. NACHTSHEIM

MURALI S. SHANKER

Curtis L. Carlson School of Management
University of Minnesota
Minneapolis, MN 55455

Abstract: Field service managers are often faced with the problem of balancing the number of technicians, territory size, and field service quality. This paper presents an approximate state-dependent queuing model that can help field service managers make these tradeoffs. Simulation experiments over a variety of field service environments demonstrate that this model is quite accurate for predicting mean travel time and mean response time. The approximate queuing model has been imbedded in a decision support system and implemented by a Fortune 100 company. Management found the decision support system very useful in making important field service decisions.

■ Field service planning is becoming increasingly important as "high tech" machines such as computers, robots, communication systems, and copy machines are becoming more popular and more widely dispersed geographically. Field service repair for these products is a major competitive issue (Blumberg [3]). In order to compete effectively, field service managers must frequently make difficult tradeoffs between service level (typically measured by the mean response time) and service system cost (technician and travel cost).

The problem cannot be adequately addressed with any queuing theory model currently in the literature. Simulation is not recommended because of the high computational expense.

This paper presents an approximate state-dependent queuing model that can help field service managers make tradeoffs between the number of technicians, territory size (for square territories), and mean response time. The model can be used to quickly evaluate the performance of alternative territory designs to help field service managers make both tactical and operational decisions.

The first section of the paper reviews the relevant field service repair literature. The second section presents a model for determining the expected travel time given the steady state probabilities of n failures in the system, P_n. The third section presents an approximate state-dependent M/G/s queuing model for estimating the P_n and the expected steady state response time for s servers (technicians). The fourth section presents a simulation experiment which demonstrates that the approximate M/G/s model is quite accurate in a variety of field service territory environments. The paper concludes with a summary of the results and a brief overview of the implementation of the model.

Received March 1988; revised March 1989 and October 1989. Handled by the Department of Engineering Statistics and Applied Probability.

Field Service Repair Literature

The field service repair literature can be divided into two categories:

1. *Round trip travel* from a central facility to the failure location, and,

2. *Sequential trip travel* from one failure location to another.

The round trip travel models are most useful for low utilization servers (such as fire engines) where the server has a central "home base" location. The literature in this area is fairly extensive (Chelst and Jarvis [4], Chiu and Larson [5], Berman, Larson, and Chiu [2], Eilon, Watson-Gandy, and Christophides [7], Kolesar [15], Kolesar and Blum [16], Larson [17]).

Sequential trip travel models are most useful for environments with high service utilization and where the servers do not return to a central location but rather move from one failure location to another throughout the day. In this type of field service environment, when the servers (technicians) finish servicing a failure (customer), they call in to a dispatcher for their next assignment. If a repair is not complete at the end of the day, the repair is resumed at the beginning of the next workday without any additional "response time" for the customer. (In this case technicians travel on their own time and arrive at the customer location at the beginning of the next workday.)

Sequential trip models can be further broken down into finite and infinite calling population models. Agnihothri [1] deals with single server finite calling population models. He also suggests heuristic methods for handling multiple servers in a closed queuing network.

Most "high tech" field service repair systems deal with sequential trip, high machine population (i.e., infinite calling population) environments. The remainder of the literature review and the remainder of the paper, therefore, deal exclusively with sequential trip, infinite calling population models.

Dzubow [6] developed a simulation approach for the multiple server, sequential trip, infinite calling population field service system. He presents a methodology for optimizing field service decision variables using simulation as the means to evaluate system performance.

Hambleton [10] proposed a simple square root model for a multiple server, sequential trip, infinite calling population field service system:

$$E(T) = K_a(a^2 + b^2)^{1/2}/(2vs^{1/2}) \tag{1a}$$

where,

$E(T)$ = expected travel time.

a,b = dimensions of the rectangular field service territory.

v = constant travel velocity (speed).

K_a = constant used to fit the model to historical data.

s = number of servers assigned to the territory.

Hambleton assumes that machines are uniformly distributed throughout a rectangle of size ($a \times b$) and that the mean distance between two random points is one half of the diagonal distance (i.e., $(a^2+b^2)^{1/2}/2$). Kolesar and Blum [16] developed and tested a very similar "square root law" model for fire engines traveling from fixed locations to fire emergency locations:

$$E(T) = K_b(ab)^{1/2}/s^{1/2}. \tag{1b}$$

They state that this "approximation is reasonable if the probability that all units are busy is small." Note that the Hambleton and the Kolesar/Blum models are essentially equivalent in form.

For $s \geq 2$ these models are not accurate as travel time becomes a function of the number of failures in the system (state of the system). For example, if the queue of failures waiting for service is not empty, the first available server is dispatched to the first failure in the queue and the expected travel time is no better than a system with only one server. However, if no failures are in the system when a failure occurs, all s servers are available and the nearest server is dispatched to service the failure. In this case the expected travel time will be much less than that with only one server available. Expected travel time and service time, therefore, are dependent upon the state of the system and must be defined in terms of P_n, the steady state probability of n failures in the system, which is by itself a function of the expected travel time.

In order to evaluate the Hambleton and Kolesar/Blum square root models (equations 1a and 1b), we conducted an extensive simulation experiment of high utilization field service environments ($.6 \leq \varrho \leq .9$). (We define utilization ϱ as the percent of the time that the servers are busy with either travel or machine repair.) We found the best constant K_a for $s=1$ and then applied the equation (1a) to estimate the expected travel times for $s=2,3,\ldots,6$. The estimated travel times were very different from the simulation results (average absolute error $\geq$ 50%). We conclude, therefore, that the multiple server sequential trip field service support problem cannot be adequately modeled with a square root model.

Smith [21] presented a queueing model for the single server, sequential trip, infinite calling population field service system:

$$E(T) = (a+b)/(3v), \tag{2}$$

$$V(T) = c(a^2+b^2)/(18v^2) \tag{3}$$

where,

a,b = dimensions of the rectangular field service territory.

v = constant travel velocity (speed).

$E(T)$ = expected travel time.

$V(T)$ = variance of travel time.

c = constant used to correct for the slight autocorrelation in travel times caused by the fact that the ending point of one trip is the beginning point of the next trip. (Smith suggests that c is approximately 1.1 for a square territory.)

This model assumes that all travel is along the x and y dimensions of a rectangle (L_1 metric) with dimensions (a and b) and that failures are distributed randomly with uniform density throughout the rectangular territory. Smith combines the above travel time model with the Pollaczek-Khinchine model [12] to determine the mean response time for a single server in a rectangular territory.

The time line in Figure 1 defines several important terms that are used in Smith's M/G/1 model and in this paper. Service time is defined as the sum of travel time and machine repair time. Response time is the sum of queue time and travel time.

Given that the system is a single channel queue with Poisson arrivals of failures with mean rate λ and an arbitrary machine repair time distribution with known mean, τ_m, and standard deviation, σ_m, Smith applies the Pollaczek-Khinchine M/G/1 model to calculate the expected queue time:

$$W_q = \varrho\tau(1+\sigma^2/\tau^2)/[2(1-\varrho)] \tag{4}$$

where,

τ = mean service time
= mean travel time $E(T)$ plus mean time to repair τ_m
= $(a+b)/(3v) + \tau_m$,

ϱ = utilization
= mean failure rate/mean service rate
= $\lambda\tau$,

σ^2 = service time variance
= travel time variance $V(T)$ plus time to repair variance σ_m^2
= $c(a^2+b^2)/(18v^2) + \sigma_m^2$.

Expected response time $E(R)$ is then:

$$E(R) = W_q + E(T) = \varrho\tau(1+\sigma^2/\tau^2)/[2(1-\varrho)] + (a+b)/(3v). \tag{5}$$

Smith points out that ϱ, τ, and σ^2 depend only upon the territory size parameters a and b and known parameters τ_m, σ_m, and v. He then presents a method for optimizing the territory size parameter, a, assuming a square territory ($a=b$) for a single server given a desired expected response time $E(R)$. Smith also points out that service times are autocorrelated (the destination for one trip is the origin for the next) and argues that the error induced by assuming that service times are IID is small. The following sections present a field service planning model that extends Smith's work to the sequential trip, multiple server field service environment.

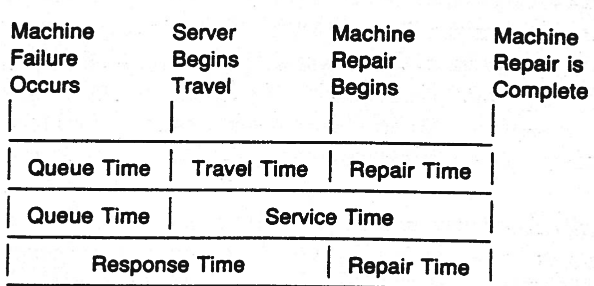

Figure 1. Time Line

A Model for the Expectation of Sequential Trip Travel Time

We seek to develop a model that expresses the expected travel time $E(T)$ as a function of the mean failure rate, λ, the mean and standard deviation of the arbitrary distribution of the time to repair, τ_m and σ_m, the number of servers, s, the dimensions of the rectangular territory, $(a \times b)$, and the constant travel rate, v. We will assume, as previous authors have, that either the L_1 or L_2 metric can adequately reflect the travel times within the territory and that failures are uniformly distributed within the territory. The L_p metric is superior to either the L_1 or L_2 metrics Love and Morris [19] [20]. If desired our results can be generalized to the L_p space, but the additional accuracy may not be worth the additional complexity of such an approach. We also assume as did Smith [21] that the autocorrelation in the travel times is not significant.

As discussed above, if all s servers are busy when a failure occurs, the first server that becomes available is dispatched to the failure. Expected travel time for this case is the same as expected travel time with only one server. If exactly one server is available at the time a failure occurs, the expected travel time is again the same as the one server case. If two or more servers are available, the closest server is dispatched and the expected travel time is significantly less than in the one server case.

We can develop a model for the expected travel time given that k servers are available. Suppose that a failure occurs at some location (x_0, y_0) in the $(a \times b)$ rectangular territory and that k servers are available at random locations $[(x_i, y_i), i=1,\ldots,k]$. As stated above, if $k=0$, the first server that becomes available will be dispatched. Let (x_i, y_i) and t_i denote the location and travel times of the i-th available server $(1 \le i \le k)$ to the failure location. Since the closest available server will be dispatched, actual travel time to service a failure will be:

$$T_k = \min_{1 \le i \le k} (t_i) \tag{6}$$

where,

T_k = travel time from closest server to the failure with k servers available.

t_i = travel time from server i to the failure location.
= $\{|x_i - x_0|^d + |y_i - y_0|^d\}^{1/d}/v$ for the L_d metric.

The density, expectation, and variance for the random variable T_k for the L_1 metric are derived by Hill and Nachtsheim [11] for general values of a and b. The expressions are lengthy, explicit functions of both a and b. (Larson and Odoni [18] page 142 define a similar problem but do not provide expressions for the parameters of interest.) Often in practice, however, the territory can be assumed to be square ($a=b$) which results in a much simpler form. In what follows we restrict attention to this special case. For $a=b$ and $d=1$, the density of T_k, derived in the Appendix, is given by:

$$f_{T_k}(T_k) = \begin{cases} kv\{1-(v^4t^4-8av^3t^3+12a^2v^2t^2)/(6a^4)\}^{k-1} \\ \times\ (4v^3t^3-24a^2t^2+24a^2vt)/6a^4; \\ \qquad\qquad\qquad\qquad\qquad (\text{for } 0\leq t\leq a). \\ kv\{(v^4t^4-8av^3t^3+24a^2v^2t^2-32a^3vt+10a^4)/(6a^4)+1\}^{k-1} \\ \times\ (-4v^3t^3+24av^2t^2-48a^2vt+32a^3)/6a^4; \\ \qquad\qquad\qquad\qquad\qquad (\text{for } a<t\leq 2a). \end{cases} \quad (7)$$

Analytic expressions for the mean and variance can be obtained for any k by direct integration of (7) (Hill and Nachtsheim [11]). Results for $1\leq k\leq 6$ are summarized for $a=b=v=1$ in Table 1. These $E_0(T_k)$ and $V_0(T_k)$ values can be rescaled for general values of a and v to determine $E(T_k)$ and $V(T_k)$. For $1\leq k\leq s$ and for general values of a and v, $E(T_k)=E_0(T_k)a/v$ and $V(T_k)=V_0(T_k)a^2/v^2$. The proposed model, therefore, can be implemented in territory planning optimization simply by using the parameters from Table 1.

The above model assumes that travel time is a linear function of distance. Kolesar [15] modeled travel time as a nonlinear function of distance. A change of variable can be used to yield the density of T_k when travel time is a non-linear function of distance. This extension has not been pursued here.

Expressions are very difficult to derive for the L_2 metric. Equation (6) was simulated with $d=2$, therefore, to approximate $E_0(T_k)$ and $V_0(T_k)$ for a unit square territory ($a=b=1$) with speed $v=1$. Again, these values can be rescaled for other values of a and v to calculate $E(T_k)$ and $V(T_k)$. The L_2 simulation estimates have relative precision better than 0.001 for a 90 percent confidence interval on the mean (sample sizes were in excess of 1 million observations). Table 1 can be used to determine values for $E(T_k)$ and $V(T_k)$ for a square territory for up to six servers for either the L_1 or the L_2 metric.

Table 1. $E_0(T_k)$ and $V_0(T_k)$ for Square Territories with the L_1 and L_2 Metrics

Number of Servers Available	L_1 Metric		L_2 Metric	
k	$E_0(T_k)$	$V_0(T_k)$	$E_0(T_k)$	$V_0(T_k)$
1	.6666667	.1111111	.521171	.061441
2	.4776000	.0616676	.389067	.042323
3	.3902838	.0426040	.322003	.031664
4	.3373742	.0324106	.279158	.024905
5	.3010032	.0260590	.249567	.020399
6	.2740734	.0217311	.227051	.017094

We note that for $n\geq s$ that $E(T_{s-n})=E(T_1)$ and calculate the expected travel time for the territory with s servers as:

$$\begin{aligned} E(T) &= \sum_{n=0}^{\infty} P_n E(T_{s-n}) \\ &= \sum_{n=0}^{s-1} P_n E(T_{s-n}) + \sum_{n=s}^{\infty} P_n E(T_{s-n}) \\ &= \sum_{n=0}^{s-1} P_n E(T_{s-n}) + E(T_1)\sum_{n=s}^{\infty} P_n \\ &= \sum_{n=0}^{s-1} P_n E(T_{s-n}) + (1-\sum_{n=0}^{s-1} P_n)\, E(T_1) \\ &= \sum_{n=0}^{s-1} P_n E_0(T_{s-n})a/v + (1-\sum_{n=0}^{s-1} P_n)\, E_0(T_1)a/v. \end{aligned} \quad (8)$$

A Model for Expected Steady State Response Time with Multiple Servers

Equation (8) is of little use for field service planning, however, unless we have a means to estimate the steady state probabilities P_n. In this section we will present an approximate model for calculating these probabilities and for estimating the expected response time $E(R)$ for a multiple server territory.

If service time (the sum of travel time and time to repair) is assumed to be exponentially distributed and not state-dependent, and if we knew the expected travel time, $E(T)$, we could use the standard M/M/s model (Hillier and Lieberman [12]) with $\mu=[E(T)+\tau_m]^{-1}$, $\varrho=\lambda/(s\mu)$, and $\theta_s=(\lambda/\mu)^s/s!$ to estimate the expected time in queue as:

$$W_q = P_0\theta_s\varrho/[\lambda(1-\varrho)^2]. \quad (9)$$

We could then calculate expected response time as:

$$E(R) = W_q + E(T) = P_0\theta_s\varrho/[\lambda(1-\varrho)^2] + E(T). \quad (10)$$

However, the M/M/s queuing model is inappropriate for field service planning as the travel time component of service times is a function of the state of the system and has an expectation defined by equation (8).

We can modify the M/M/s model so that it has state-dependent service times by defining the state-dependent service rates as:

$$\mu_n = \begin{cases} n[E(T_{s-n+1}) + \tau_m]^{-1} & \text{for } n\leq s \\ s[E(T_1) + \tau_m]^{-1} & \text{for } n > s \end{cases} \quad (11)$$

where μ_n is the rate at which all busy servers achieve service completions. The state-dependent M/M/s model is different from the M/M/s model in that:

$$\theta_n = \prod_{k=1}^{n}(\lambda/\mu_k) \tag{12}$$

which becomes:

$$\theta_n = \begin{cases} (\lambda^n/n!)\prod_{k=1}^{n}[E(T_{s-k+1}) + \tau_m] & \text{for } n \le s \\ \\ (\lambda/s)^{n-s}[E(T_1) + \tau_m]^{n-s}\,\theta_s & \text{for } n > s \end{cases}, \tag{13}$$

$$\varrho_{max} = \lambda\,[E(T_1) + \tau_m], \tag{14}$$

$$P_0 = \left\{\sum_{n=0}^{\infty}\theta_n\right\}^{-1} = \left\{\sum_{n=0}^{s-1}\theta_n + \theta_s/(1-\varrho_{max})\right\}^{-1}, \tag{15}$$

$$W_q = P_0\theta_s\varrho_{max}/[\lambda(1-\varrho_{max})^2]. \tag{16}$$

This state-dependent model is unlike the state-dependent model in Hillier and Lieberman [12] in that it has a closed-form expression for the infinite series in equation (20) and, therefore, a closed form expression for W_q and $E(R)$. The other M/M/s equations are still applicable. This model does not require that we know $E(T)$ *a priori*. We need know only the $E(T_k)=E_0(T_k)a/v$ (where the constants $E_0(T_k)$ are from Table 1).

Unfortunately, this state-dependent M/M/s model is not adequate for field service planning as service time is the sum of two random variables (travel time and time to repair) and therefore, is not exponential. In a large set of simulation experiments this model overestimated expected response time by 20 to 65 percent. (Green and Kolesar [9] found that the police service times did fit an exponential; however, this is significantly different context.)

In order to handle this difficulty, we need to develop a state-dependent queuing model with general service times. Tijms [22] and others have presented approximate M/G/s models. Federguen and Tijms [8] have developed a recursive model for the M/G/s queue with a variable service rate. In our opinion, none of these models could be modified as easily as Yao's M/G/s model [23] to handle state-dependent service times. Yao's M/G/s B-model is:

$$\varrho = \lambda/(s\mu), \tag{17}$$

$$r_1 = 2(\lambda-\mu)/(\lambda+\mu^3\sigma^2), \tag{18}$$

$$r_s = 2(\lambda-s\mu)/(\lambda+s\mu^3\sigma^2), \tag{19}$$

$$\theta_n = (\lambda/\mu)^n/n!, \tag{20}$$

$$P_0 = \left\{\sum_{n=0}^{s-1}\theta_n+\theta_s/(1-\varrho) + (s\varrho/r_1)[\exp(r_1/2)-\exp(-r_1/2)-r_1]\right\}^{-1}, \tag{21}$$

$$P_1 = P_0(\theta_1/r_1)[\exp(r_1/2)-\exp(-r_1/2)], \tag{22}$$

$$P_n = P_0\theta_n \quad \text{for } n = 2,\ldots,s, \tag{23}$$

$$W_q = P_0\theta_s\varrho/[\lambda(1-\varrho)(1-\exp(r_s))]. \tag{24}$$

In order to make Yao's model a state-dependent M/G/s model for field service planning, we note that the Yao model equations (25)-(29) are similar to M/M/s equations (11)-(14). We can make the Yao model state-dependent as we did with the M/M/s model by making the θ_n state-dependent with equation (13) and substituting ϱ_{max} from equation (14) for ϱ. r_1 and r_s are also redefined with state-dependent service times and variances. For r_1 we define $\mu=[E(T_s)+\tau_m]^{-1}$ and $\sigma^2=V(T_s)+\sigma_m^2$ and for r_s we define $\mu=[E(T_1)+\tau_m]^{-1}$ and $\sigma^2=V(T_1)+\sigma_m^2$.

In summary, the proposed approximate state-dependent M/G/s model for field service planning is:

$$r_1 = \frac{2\{\lambda - [E_0(T_s)a/v + \tau_m]^{-1}\}}{\lambda + [E_0(T_s)a/v + \tau_m]^{-3}\,[V_0(T_s)a^2/v^2 + \sigma_m^2]}, \tag{25}$$

$$r_s = \frac{2\{\lambda - s[E_0(T_1)a/v + \tau_m]^{-1}\}}{\lambda + s[E_0(T_1)a/v + \tau_m]^{-3}\,[V_0(T_1)a^2/v^2 + \sigma_m^2]}, \tag{26}$$

$$\theta_n = \begin{cases} (\lambda^n/n!)\prod_{k=1}^{n}[E_0(T_{s-k+1})a/v + \tau_m] & \text{for } n \le s \\ \\ (\lambda/s)^{n-s}\,[E_0(T_1)a/v + \tau_m]^{n-s}\,\theta_s & \text{for } n \ge s \end{cases}, \tag{27}$$

$$\varrho_{max} = \lambda[E_0(T_1)a/v + \tau_m], \tag{28}$$

$$P_0 = \left\{\sum_{n=0}^{s-1}\theta_n + \theta_s/(1-\varrho_{max}) + s\varrho_{max}/r_1\,[\exp(r_1/2)-\exp(-r_1/2)-r_1]\right\}^{-1}, \tag{29}$$

$$P_1 = P_0\theta_1/r_1[\exp(r_1/2)-\exp(-r_1/2)], \tag{30}$$

$$P_n = P_0\theta_n \quad \text{for } n = 2,\ldots,s, \tag{31}$$

$$E(T) = \sum_{n=0}^{s-1}P_nE_0(T_{s-n})(a/v) + (1 - \sum_{n=0}^{s-1}P_n)E_0(T_1)(a/v), \tag{32}$$

$$E(R) = W_q + E(T) = P_0\theta_s\varrho_{max}/[\lambda(1-\varrho_{max})(1-\exp(r_s))] + E(T). \tag{33}$$

The expected response time $E(R)$, therefore, is a function of the constants $E_0(T_k)$ and $V_0(T_k)$, the system parameters λ, τ_m, σ_m, v, and the decision variables s and a. The model can be used to explore tradeoffs between the expected response time $E(R)$, and the decision variables s and a.

Simulation Experiment

Since the above model is approximate, it is important that we test the model in a variety of field service repair environments. We developed a simulation, therefore, of a field service support system with Poisson arrivals of failures at failure locations (x,y) where x and y are uniformly distributed on the interval $[0,a]$ The simulation results can then be compared to the expected values estimated by the approximate model.

The field service problem environments for this experiment are characterized by:

1. Number of servers, s, set at 1, 2, 3, 4, 5, and 6. (Few field service territories in our experience have more than 6 servers.)

2. Target utilization set at .6, .7, .8, and .9. (These are roughly the same limits set by Smith [21].)

3. Poisson arrivals of failures to the system with failure rates λ are adjusted to achieve the target utilizations with $\lambda = s\hat{\varrho}/[E(T)+\tau_m]$.

4. Square field service territory with $a=50$ miles, constant speed $v=30$ miles per hour, and distance defined by the L_1 metric.

5. The time to repair distribution is exponential with $\tau_m = \sigma_m = 2$ hours.

As suggested by Kelton and Law [14], each simulation run is initialized with the expected number of failures in the queue, λW_q, where W_q is estimated using the approximate state-dependent M/G/s model. The batch means approach was used to provide serially independent observations. Each field service problem environment is simulated until a 90 percent confidence interval on the mean response time has a relative precision better than 0.01.

Table 2 summarizes the problem environments in this experiment. Actual utilizations are slightly different from target utilizations because we do not have an exact method for determining the arrival rate required to achieve a given target utilization for $s>1$. The long simulation run lengths support our earlier contention that simulation is not an adequate approach for this problem if the field service planner must simulate many field service environments.

Table 3 reports the approximate queuing model expected travel times, $E(T)$, as well as the mean travel times from the simulation experiments for the 20 problem environments. Table 4 reports the same information expressed as percent differences from the simulated values where:

$$\text{Percent Difference} = 100\,[\text{Simulation Mean} - E(T)]/\text{Simulation Mean}. \tag{34}$$

The mean percent difference across all 20 problem environments in expected travel time is 1.3 percent. The mean absolute percent difference in expected travel time is also 1.3

Table 2. Simulation Experiment Problem Environments

Number of Servers s		Target Utilization .6	.7	.8	.9
1	Mean Arrival Rate λ	.19286	.22500	.25714	.28929
	Number of Failures	345,000	535,000	770,000	1,010,000
	Simulation Utilization	.60014	.70022	.80024	.90030
2	Mean Arrival Rate λ	.39104	.45175	.51236	.57169
	Number of Failures	310,000	500,000	750,000	980,000
	Simulation Utilization	.59359	.69122	.78878	.88457
3	Mean Arrival Rate λ	.60124	.69053	.77838	.86725
	Number of Failures	285,000	485,000	740,000	995,000
	Simulation Utilization	.59423	.69388	.79039	.89016
4	Mean Arrival Rate λ	.81881	.93949	1.05306	1.16693
	Number of Failures	265,000	485,000	740,000	990,000
	Simulation Utilization	.59644	.69803	.79496	.89438
5	Mean Arrival Rate λ	1.04975	1.19136	1.33738	1.44995
	Number of Failures	260,000	455,000	745,000	950,000
	Simulation Utilization	0.60238	0.69776	0.80134	0.88318
6	Mean Arrival Rate λ	1.28019	1.45007	1.59265	1.77921
	Number of Failures	255,000	475,000	700,000	1,025,000
	Simulation Utilization	0.60316	0.70031	0.78549	0.90311

Number of Servers s	Model	Target Utilization .6	.7	.8	.9
1	Queuing Model $E(T)$[a]	1.11111	1.11111	1.11111	1.11111
	Simulation Mean	1.11145	1.11184	1.11111	1.11100
2	Queuing Model $E(T)$	1.02801	1.05114	1.07206	1.09073
	Simulation Mean	1.03759	1.05901	1.07776	1.09431
3	Queuing Model $E(T)$	0.95417	0.99581	1.03473	1.07220
	Simulation Mean	0.96729	1.00920	1.04549	1.07852
4	Queuing Model $E(T)$	0.89025	0.94856	1.00291	1.05649
	Simulation Mean	0.91196	0.96866	1.01838	1.06506
5	Queuing Model $E(T)$	0.83855	0.90598	0.97743	1.03318
	Simulation Mean	0.86521	0.92877	0.99490	1.04439
6	Queuing Model $E(T)$	0.79111	0.86909	0.93849	1.03320
	Simulation Mean	0.82120	0.89541	0.95930	1.04459

Table 3. Approximate Queuing Model Expected Travel Time and Simulation Model Mean Travel Time

Note: [a]$E(T)$ is exact for $s=1$.

Table 4. Percent Difference Between Approximate Queuing Model Expected Travel Time and Simulation Model Mean Travel Time

Number of Servers s	Target Utilization .6	.7	.8	.9
1[a]	0.03%	0.07%	0.00%	0.01%
2	0.92%	0.74%	0.53%	0.33%
3	1.36%	1.33%	1.03%	0.59%
4	2.38%	2.08%	1.52%	0.80%
5	3.08%	2.45%	1.76%	1.07%
6	3.66%	2.94%	2.17%	1.09%

Note: [a]$E(T)$ is exact for $s=1$.

percent.

Table 5 reports the approximate queuing model expected response times $E(R)=W_q+E(T)$ as well as the mean response times from the simulation experiments for all 20 problem environments. The Pollaczek-Khinchine M/G/1 results (Smith's model) are also shown for $s=1$. These results show that the proposed approximate model does very well even for the single server case. Smith's travel time variance correction factor $c=1.1$ does not make a noticeable improvement.

Table 6 expresses the approximate queuing model expected response time $E(R)$ in terms of percent differences from the simulation mean response times. The mean percent difference in expected response time is only 0.02 percent and the mean absolute percent difference is .88 percent. The worst case absolute percent error (3 percent) is for the single server model.

Conclusions

This paper has presented an approximate state-dependent

Table 5. Approximate Queuing Model Expected Response Time and Simulation Model Mean Response Time

Number of Servers s	Model	Target Utilization .6	.7	.8	.9
1	Queuing Model $E(R)$	4.58058	6.43830	10.16345	21.37624
	Simulation Mean	4.44734	6.33775	10.10821	21.51625
	M/G/1 Model ($c=1.0$)[a]	4.48313	6.35648	10.10317	21.34325
	M/G/1 Model ($c=1.1$)[b]	4.49058	6.36806	10.12302	21.38790
2	Queuing Model $E(R)$	2.35559	3.23791	4.96078	9.53988
	Simulation Mean	2.30910	3.20083	4.97528	9.59303
3	Queuing Model $E(R)$	1.69090	2.30965	3.52462	7.04952
	Simulation Mean	1.66588	2.30203	3.53758	7.10791
4	Queuing Model $E(R)$	1.36409	1.87575	2.86732	5.81606
	Simulation Mean	1.36169	1.88591	2.88709	5.84481
5	Queuing Model $E(R)$	1.18229	1.60154	2.51907	4.41901
	Simulation Mean	1.19251	1.60356	2.53329	4.42717
6	Queuing Model $E(R)$	1.04621	1.42295	2.04863	4.71568
	Simulation Mean	1.06382	1.44032	2.06257	4.77110

Notes: [a]Smith's M/G/1 model with $c=1.0$. [b]Smith's M/G/1 model with $c=1.1$.

Table 6. Percent Difference Between Approximate Queuing Model Expected Response Time and Simulation Model Mean Response Time				
Number of Servers	Target Utilization			
s	.6	.7	.8	.9
1	−3.00%	−1.59%	−0.55%	0.65%
2	−2.01%	−1.16%	0.29%	0.55%
3	−1.50%	−0.33%	0.37%	0.82%
4	−0.18%	0.54%	0.68%	0.49%
5	0.86%	0.13%	0.56%	0.18%
6	1.66%	1.21%	0.68%	1.16%

M/G/s model for estimating steady state expected travel times $E(T)$ and expected response times $E(R)$ for a field service territory defined by the system parameters: λ, the mean failure rate, τ_m and σ_m, the mean and standard deviation of the arbitrary time to repair distribution, v, the constant travel rate (speed), and the decision variables: s, the number of servers (technicians), and a, the square territory size parameter. The model can be applied with either the L_1 or L_2 metric. The paper also suggests ways to extend the model to non-square territories.

Simulation experiments have demonstrated that the approximate model is able to make accurate estimates for both the expected travel time and expected response time in a variety of field service repair environments. Mean absolute percent differences between the queuing model expected values and the simulation means are on the order of 1 percent for these experiments.

The model can help field service planners to quickly evaluate tradeoffs between field service workforce size, territory size, and mean response time. The model was implemented in a decision support system by a Fortune 100 company to facilitate the analysis of a major change in workforce levels. The DSS was designed so that the management could gain insight into the tradeoffs between response times, expected travel times, and workforce levels. The model was useful in convincing management that small reductions in the number of technicians (servers) would have a significant effect on total system performance. Decreasing the number of technicians simultaneously increases travel time and reduces overall service capacity and results in significantly poorer service.

Acknowledgements

This research was supported in part by grants from the University of Minnesota Operations Management Center and the IBM Corporation.

REFERENCES

[1] Agnihothri, Saligrama R., *Performance Evaluation of Field Service Territories*, Ph.D. Dissertation, University of Rochester (1985).

[2] Berman, O., Larson, R. C., and Chiu, S. S., "Optimal Server Location on a Network Operating as an M/G/1 Queue," *Operations Research*, 33, 746-771 (1985).

[3] Blumberg, Donald F., "Management Systems for Field Service Productivity Improvement," *Omega*, 9, 419-428 (1981).

[4] Chelst, Kenneth and Jarvis, James P., "Estimating the Probability Distribution of Travel Times for Urban Emergency Service Systems," *Operations Research*, 27, 199-205 (1979).

[5] Chiu, Samuel S., and Larson, Richard C., "Locating an n-Server Facility in a Stochastic Environment," *Computers & Operations Research*, 12, 6, 509-516 (1985).

[6] Dzubow, Steven R., *A Methodology for the Design and Control of Logistic Service Systems*, Unpublished Doctoral Thesis, University of Pennsylvania (1972).

[7] Eilon, Samuel, Watson-Gandy, C. D. T., and Christophides, Nicos, *Distribution Management: Mathematical Modelling and Practical Analysis*, Hafner Publishing Company, New York (1971).

[8] Federgruen, A., Tijms, H. C., "Computation of the Stationary Distribution of the Queue Size in an M/G/1 Queueing System with Variable Service Rate," *Journal of Applied Probability*, 17, 515-522 (1980).

[9] Green, Linda, and Kolesar, Peter, "The Feasibility of One-Officer Patrol in New York City," *Management Science*, 30, 8, 964-981 (1984).

[10] Hambleton, R. S., "A Manpower Planning Model for Mobile Repairman," *Journal of Operational Research Society*, 33, 621-627 (1982).

[11] Hill, Arthur V., and Nachtsheim, Christopher J., "An Approach for Finding the Conditional Expectation and Variance of Travel Times in an L_1 Rectangular Territory with k Servers Available," Operations Management Working Paper, School of Management, University of Minnesota (1986).

[12] Hillier, Frederick S., and Lieberman, Gerald J., *Operations Research*, Fourth Edition, Holden-Day, San Francisco (1986).

[13] Johnson, Norman L., and Kotz, Samuel, *Continuous Univariate Distributions-1*, John Wiley & Sons, New York (1970).

[14] Kelton, W. David, and Law, Averill M., "The Transient Behavior of the M/M/s Queue, with Implications for Steady-State Simulation," *Operations Research*, 33, 378-396 (1985).

[15] Kolesar, Peter, "Determining the Relation between Fire Engine Travel Times and Travel Distances in New York City," *Operations Research*, 23, 603-613 (1975).

[16] Kolesar, Peter, and Blum, Edward H., "Square Root Laws for Fire Engine Response Distances," *Management Science*, 19, 12, 1368-1378 (1973).

[17] Larson, Richard C., "Approximating the Performance of Urban Emergency Systems," *Operations Research*, 23, 845-868 (1979).

[18] Larson, Richard C., and Odoni, Amadeo R., *Urban Operations Research*, Prentice Hall, Englewood Cliffs, NJ (1981).

[19] Love, Robert F., and Morris, J. G., "Modelling Inter-City Road Distances by Mathematical Functions," *Operational Research Quarterly*, 23, 1, 61 (1972).

[20] Love, Robert F., and Morris, J. G., "Mathematical Models of Road Travel Distances," *Management Science*, 25, 2, 130-139 (1979).

[21] Smith, Stephen A., "Estimating Service Territory Size," *Management Science*, 25, 301-311 (1979).

[22] Tijms, Henk C., *Stochastic Modeling and Analysis — A Computational Approach*, John Wiley & Sons, New York (1986).

[23] Yao, David D., "Refining the Diffusion Approximation for the M/G/m Queue," *Operations Research*, 33, 1266-1277 (1985).

Appendix

Distribution of Sequential Travel Times in an L_1 Distance Metric Rectangular Territory with k Servers Available

The L_1 distance, D, from any server at location (x_i,y_i) to a random failure at location (x_0,y_0) is $D=X+Y$, where $X=|x_i-x_0|$ and $Y=|y_i-y_0|$. The corresponding travel time is

D/v. By assumption, x_i and x_0 are independent, uniform random variables on $[0,a]$. Similarly, y_i and y_0 are independent and uniform on $[0,b]$. In what follows we take $a \leq b$ without loss of generality. If U_1 and U_2 are independent uniform random variables on $[0,1]$, it is easy to show that the density of $Z=|U_1-U_2|$ is $f_Z(z)=2(1-z)$ for $0 \leq z \leq 1$.

A change of variable yields the densities of X and Y, $f_X(x)=(2/a)(1-x/a)$ and $f_Y(y)=(2/b)(1-y/b)$. Convolving X and Y to find the cumulative distribution function of D, we have by independence:

$$F_D(d) = \iint_{A_d} f_X(x) f_Y(y)\,dx\,dy$$

where

$$A_d = \{(x,y) \mid 0 \leq x \leq a;\ 0 \leq y \leq b;\ x+y \leq d\}.$$

Integration yields:

$$F_D(d) = \begin{cases} (d^4-(4a+4b)d^3+12abd^2)/6a^2b^2; & (0 \leq d \leq a) \\ (5b^3+8ab^2-8b^2d)/6a^2b \\ \quad -(d^2-2(a+b)d+b^2+2ab)/a^2; & (a<d \leq b) \\ \{-d^4+4(a+b)d^3-(6b^2+12ab+6a^2)d^2 \\ \quad +(4b^3+12ab^2+12a^2b+4a^3)d \\ \quad -b^4-4ab^3-4a^3b-a^4\}/6a^2b^2; & (b<d \leq a+b). \end{cases}$$

From Johnson and Kotz [13] we know that the cumulative distribution function of the minimum distance among k servers, D_k, $1 \leq k \leq s$, is:

$$F_{D_k}(d_k)=1-(1-F_D(d))^k.$$

The density of D_k follows by differentiation with respect to d. Since $T_k=D_k/v$, a change of variable yields the density of T_k, and (7) results when $a=b$. Means and variances of T_k are found by direct (symbolic) integration of (7).

Arthur V. Hill is an Associate Professor in the Operations and Management Science Department of the Curtis L. Carlson School of Management at the University of Minnesota. He received a B.A. in Mathematics from Indiana University and an M.S.I.A. and Ph.D. from the Krannert Graduate School at Purdue University. He is an associate editor for *Decision Science* and serves on the editorial boards for the *Journal of Operations Management*, the *Production and Inventory Management Journal*, and the *Journal of the Production and Operations Management Society*. Professor Hill's current research interests are in the areas of "finite scheduling" for manufacturing planning and control systems and in field service operations management.

Salvatore T. March is an Associate Professor in the Information and Decision Sciences Department in the Curtis L. Carlson School of Management at the University of Minnesota. He received his B.S., M.S., and Ph.D. degrees in Operations Research from Cornell University. Dr. March is active in teaching, research, and consulting in the area of Database Design and Information Resource Management. He has published widely in the database literature including such journals as: *ACM Transactions on Database Systems*, *ACM Computing Surveys*, *Information and Management*, *The Journal of MIS*, and *Management Science*. He is currently the Editor-in-Chief of *ACM Computing Surveys*.

Christopher J. Nachtsheim is an Associate Professor in the Operations and Management Science Department in the Curtis L. Carlson School of Management at the University of Minnesota. He received a B.A. in Mathematics from the College of St. Thomas, an M.S. in Operations Research and Statistics from Rensselaer Polytechnic Institute, and a Ph.D. in Operations Research from the University of Minnesota. His current research interests include experimental design and analysis, statistical simulation and expert systems for statistical computing. He is currently an associate editor of the *Journal of the American Statistical Association*, and the *Journal of Statistical Simulation and Computation*. He is also a Department Editor for the *Journal of Quality Technology*. He is a member of ASA, ASQC, IMS and RSS.

Murali S. Shanker is a doctoral candidate in the Operations and Management Science Department in the Curtis L. Carlson School of Management at the University of Minnesota. He received a B.S. in mathematics and M.S. in Operations Research from the University of Madras, India. He is a member of TIMS, ORSA, IEEE, and DSI. His thesis topic is dynamic task allocation for distributed simulations.