PRODUCTION AND OPERATIONS MANAGEMENT
POMS
Vol. 15, No. 1, Spring 2006, pp. 103–116
ISSN 1059-1478 | 06 | 1501 | 103$1.25
© 2006 Production and Operations Management Society

Modeling and Managing the Percentage of Satisfied Customers in Hidden and Revealed Waiting Line Systems

Chester Chambers • Panagiotis Kouvelis
Cox School of Business, Southern Methodist University, Dallas, Texas 75275-0333, USA
School of Business, Washington University, St. Louis, Missouri 63130-4899, USA
cchamber@mail.cox.smu.edu • Kouvelis@olin.wustl.edu

We perform an analysis of various queueing systems with an emphasis on estimating a single performance metric. This metric is defined to be the percentage of customers whose actual waiting time was less than their individual waiting time threshold. We label this metric the Percentage of Satisfied Customers (PSC.) This threshold is a reflection of the customers' expectation of a reasonable waiting time in the system given its current state. Cases in which no system state information is available to the customer are referred to as "hidden queues." For such systems, the waiting time threshold is independent of the length of the waiting line, and it is randomly drawn from a distribution of threshold values for the customer population. The literature generally assumes that such thresholds are exponentially distributed. For these cases, we derive closed form expressions for our performance metric for a variety of possible service time distributions. We also relax this assumption for cases where service times are exponential and derive closed form results for a large class of threshold distributions. We analyze such queues for both single and multi-server systems. We refer to cases in which customers may observe the length of the line as "revealed" queues." We perform a parallel analysis for both single and multi-server revealed queues. The chief distinction is that for these cases, customers may develop threshold values that are dependent upon the number of customers in the system upon their arrival. The new perspective this paper brings to the modeling of the performance of waiting line systems allows us to rethink and suggest ways to enhance the effectiveness of various managerial options for improving the service quality and customer satisfaction of waiting line systems. We conclude with many useful insights on ways to improve customer satisfaction in waiting line situations that follow directly from our analysis.

Key words: waiting lines; customer satisfaction; management of customer queues
Submissions and Acceptance: Received December 2003; revision received December 2004, August 2005, October 2005; accepted December 2005 by Kalyan Singhal.

1. Introduction

Many settings exist in which customers join a single-channel queue with a high level of commitment to reaching the service provider at the end of the line. For example, we may join a queue to check in at an airport with tickets in hand, or to receive a critical medical treatment, or to handle a particularly pressing personal or business matter. In such settings, each customer brings with him/her some tolerance for the wait that is to take place. If customers are polled after the experience and asked questions such as, 'Was the waiting time excessive?', some customers will answer in the affirmative while others answer in the negative even if they all experience waits of identical lengths. In this work, we will assume that each customer brings with him/her some threshold value (X) such that, as long as the actual wait (W) is less than this value ($W - X \leq 0$), he would respond that the wait was not excessive. For ease of exposition, we label all customers with $W - X \leq 0$ as satisfied and all customers with $W - X > 0$ as dissatisfied. We label the portion of customers for which $W - X \leq 0$ as the Percentage of Satisfied Customers (PSC.)

A large body of literature focuses upon objective measurements such as average waiting time or the likelihood of experiencing a wait. Such measurements are most useful when compared to some benchmark. Our approach addresses this by comparing the waiting time to customer specific thresholds. While our binary classification scheme is a dramatic simplification of reality, our approach remains valuable for several reasons. First, our metric provides a simple, global measurement of system performance that can easily be conveyed, and explained. Second, our metric generates values that may easily be verified in practice. Data from exit surveys or other customer responses allow us to classify customers as satisfied/not satisfied, while estimating the 'degree of dissatisfaction' as a function of the waiting time introduces measurement difficulties that may be impossible to overcome. Third, the analytical statement of our metric allows managers to quickly estimate the impact of proposed system changes.

1.1. Waiting Thresholds

The notion of waiting time thresholds is common in the literature on call centers (Whitt 1999a,b). In that setting, this value is interpreted as the time after which a customer will hang up after being put on hold. This behavior is often labeled 'reneging.' In our models, we typically assume that there is some characteristic of the service that minimizes this behavior. (The appendix does include limited results which include balking and reneging for M/M/c queues.) Our notion of a waiting time threshold is better described as the length of time a customer will wait, prior to the initiation of service, before becoming dissatisfied with the service provider. Our focus on the time prior to the initiation of service is consistent with the arguments of Maister (1985) who suggested that these "pre-process" waits are often (but not always) more unpleasant than in-process waits. This argument is further supported by empirical evidence (see Dube-Rioux, Schmitt, and LeClerc 1988 and the references therein.)

In many cases, customer satisfaction (or dissatisfaction) relates to the gap between customers' expectations of the performance of the service system and its actual performance (see Maister 1985 and Boulding et al. 1993 for empirical support.) This implies that customer threshold values are likely to be correlated with their expectations. This suggests that management of these expectations should have a significant impact on the PSC. On the other hand, it is also clear that these threshold values are not fully determined by customer expectations. One may be dissatisfied with a wait even when it is exactly as long as was expected.

1.2. Selected Literature

Most of the research on queueing has dealt with the mathematical theory of waiting lines, and descriptions of waiting time distributions have been developed for virtually any setting. (For recent textbooks, see Wolff 1989; Hall 1991; Gross and Harris 1998.) Several researchers have expanded this large body of work by focusing upon customer satisfaction measures which seek to include customer-specific attributes in their calculation. (For examples, see Carmon et al. 1993; Green and Kolesar 1987, 1988; Whitt 1992a,b.)

The work of Whitt (1999a,b) has been particularly influential in this area. These papers focus upon call centers, and all of the closed form metrics derived assume exponentially distributed inter-arrival times, service times, and threshold values. It is useful to have compact expressions for the PSC for a number of settings where we may relax at least one of these assumptions. For example, the assumption of exponential service times is not universally applicable even for call centers (see Inman 1999). For data on non-exponential service times in call centers, see Wardell et al. 2001. Our analysis will consider other service time distributions in such a way that we still obtain closed form expressions for the PSC.

The work of Whitt is also restricted to the consideration of cases in which the customer thresholds are independent of the state of the system. This is quite natural for call centers. We refer to such settings as "hidden queues" because some knowledge of the system state is hidden from the customer's information set. However, many cases exist in which the customer can easily see the number of customers ahead of him/her awaiting service. We refer to these settings as "revealed queues."

Common examples of hidden queues include amusement parks where the length of the line is hidden by bending it around corners or using staging areas, medical settings where some people in a waiting area are accompanying customers and some customers hold appointment slots even if they are not present, and many electronic systems in which customers cannot see how many jobs are ahead of them in line. Common waiting line situations that qualify as revealed queues include check-out areas in supermarkets, lines to enter unique sporting or entertainment events, lines at licensing centers, bus stops, theaters, etc.

2. Hidden and Revealed M/G/1 Queues

Let us consider a single-line queueing system with a single server. Customers arrive according to a Poisson process, and the required service time follows a general distribution function. Upon arrival each customer can be characterized as holding some threshold value for waiting in the line. Each customer's threshold will depend on a variety of unknown, customer-specific

Table 1 Notation Used Throughout This Work

λ	average arrival rate of customers per unit time
S	service time random variable
B	service time distribution function (i.e. $B(x) = P\{S \le x\}$)
M_i	i-th moment of the service time (i.e., $m_i = E[S^i]$)
μ	average service rate ($\mu = 1/E[S]$)
ρ	utilization factor (i.e., $\rho = \lambda\ E[S]$)
W	stationary waiting time random variable
h	(generalized) stationary waiting time density
H	stationary waiting time distribution function (i.e., $H(x) = P\{W \le x\}$)
L	number of customers in the system
X	customer's waiting time threshold upon arrival to the system
G	distribution function of X (i.e., $G(x) = P\{X \le x\}$)
n	index for number of customers in the system
Q_n	probability that there are n customers in the system (i.e., $q_n = P\{L = n\}$)
H_n	waiting time distribution function given that there are n customers ahead
Y_n	customer's expectation of the waiting time when (s)he observes n customers in the system
G_n	distribution function of Y_n
δ_{mn}	Kronecker's delta (i.e., $\delta_{mn} = 1$ for $m = n$, 0 otherwise)
$U(x)$	step function defined by $U(x) = 1$ if $x \ge 0$ and $U(x) = 0$ for $x < 0$.

factors. We model this by having each customer randomly drawing a threshold value out of a prespecified distribution. We assume that our queue is part of a stationary M/G/1 system. Our notation is rather standard, and is summarized in Table 1.

2.1. Hidden M/G/1 Queue With Exponential Distribution of Customer Thresholds

We label the random variable of interest as Z, defined as W-X. We want to calculate the distribution function of Z, let us call it F, and in particular F(0), which represents the probability that a randomly selected customer is satisfied. Since we assume Poisson arrivals, F(0) is also the PSC. Since the randomness in X arises from the heterogeneity of the customer population, we assume that X and W are statistically independent when performing our calculations. If the customer's expectation, X is drawn from an exponential distribution with rate θ, then the quantity of main interest can be expressed with the use of the following lemma.

Lemma 1. *Let I and J be independent non-negative random variables. I has a distribution function G_I and J is exponentially distributed with rate θ. Then $P\{I - J \le 0\} = \gamma(\theta)$, where γ is the Laplace-Stieltjes (L-S) transform of G_I.*

See Appendix for all proofs and additional mathematical details, including comments on numerical approaches for more general cases. Using Lemma 1, we have,

$$F(0) = P\{Z \le 0\} = P\{W - X \le 0\} = \hat{H}(\theta), \quad (1)$$

where $\hat{H}$ is the L-S transform of $H(x) = P\{W \le x\}$. For the stationary M/G/1 queueing model, it is known (see Kleinrock 1975) that

$$\hat{H}(s) \equiv \int_0^\infty e^{-sx} dH(x) = \frac{s(1-\rho)}{s - \lambda + \lambda\hat{B}(s)}, \quad (2)$$

where $\hat{B}(s)$ is the L-S transform of the service time distribution function. Thus, from (1) and (2), we obtain

$$F(0) = \frac{\theta(1-\rho)}{\theta - \lambda + \lambda\hat{B}(\theta)}. \quad (3)$$

When X ~ Exp(θ), we can use Equation (3) to state F(0) for cases where the L-S transform of the service time distribution is known, and can be evaluated at θ. Such expressions for a variety of service time distributions are shown in Table 2.

A variety of service time distributions are considered because delivery systems differ so widely in their structure. While exponential or Erlang service time distributions are quite common in the literature, automation or process standardization often leads to service times that are virtually deterministic. Hyperexponential service times may be present when different classes of customers need different levels of service, but the provider has no way of knowing which class a customer is in until he/she reaches the service provider. A Generalized Exponential service time is present when some portion of the population (ω) can be shifted to an alternate service provider or require service times of virtually 0. Finally, the ability to

Table 2 Percentage of Satisfied Customers in Hidden Queues When X ~ Exp(θ) and Service Times Follow Various Distributions

Distribution of service times	Percentage of satisfied customers (F(0))
Exponential (μ)	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \lambda\dfrac{\mu}{\theta+\mu}}$
Gamma (α, β)	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \left(\dfrac{\beta^{-1}}{\theta+\beta^{-1}}\right)^{-\alpha}}$
Deterministic	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \dfrac{\rho}{\theta}}.$
Erlang (k)	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \lambda\left(\dfrac{k\mu}{\theta+k\mu}\right)^k}$
Hyperexponential	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \lambda\sum_{i=1}^{m}\omega_i\left(\dfrac{\mu_i}{\theta+\mu_i}\right)}$
Generalized exponential	$\dfrac{\theta(1-\rho)}{\theta - \lambda + \lambda\omega\left(\dfrac{\mu}{\theta+\mu}\right)}$

model Gamma distributed service times is useful because it is a good fit for a wide range of settings not satisfactorily captured by simpler distributions.

2.2. Hidden M/M/1 Queue With General Distribution of Customer Thresholds

While it is common to assume that X ~ Exp(θ), Lemma 1 also allows us to deduce closed form expressions for F(0) when service times are assumed to be exponential as long as the L-S transform of the distribution from which customer threshold values are drawn is known. For this setting, Lemma 1 implies that we may express the percentage of satisfied customers as,

$$F(0) = (1 - \rho) + \rho(1 - \hat{G}(\mu - \lambda)), \tag{4}$$

where each customer's threshold is drawn from a distribution function G, and $\hat{G}(s)$ is its L-S transform. Closed form expressions for F(0) for several such cases are summarized in Table 3.

If customers are promised a'priori that the waiting time will be some specific value, say W$_0$, then, virtually all customers will become upset if the wait is longer than W$_0$, and the distribution of X approaches that of a deterministic setting. The notion of hyperexponential thresholds is applicable when considering a mixture of classes of customers interspersed in the same line. This is also useful in modeling a setting in which a portion of the population (ω_I) is "impatient" such that any wait is considered intolerable. A unique application of the Generalized Exponential distribution exists when some fraction of the customer pool may be distracted, entertained, or treated in such a way that their perceived cost of waiting is 0. Increasing this fraction (ω_p) of "patient" customers obviously leads to increases in customer satisfaction.

Table 3 Percentage of Satisfied Customers in Hidden Queues When Service Times ~ Exp(θ) and Customer Thresholds *X* Follow Various Distributions

Distribution of customer thresholds	Percentage of satisfied customers *F*(0)
Exponential (μ)	$(1-\rho)+\rho\left\{1-\dfrac{\mu-\lambda}{\theta+\mu-\lambda}\right\}$
Gamma (α, β)	$(1-\rho)+\rho\left\{1-\left(\dfrac{\theta}{\beta^{-1}-\lambda+\theta}\right)^{-\alpha}\right\}$
Deterministic	$(1-\rho)+\rho\left(1-\dfrac{1}{\theta(\mu-\lambda)}\right)$
Erlang (*k*)	$(1-\rho)+\rho\left\{1-\left(\dfrac{k\theta}{\mu-\lambda+k\theta}\right)^{k}\right\}$
Hyperexponential	$(1-\rho)+\rho\left\{1-\sum_{i=1}^{m}\omega_i\left(\dfrac{\theta_i}{\mu-\lambda+\theta_i}\right)\right\}$
Generalized exponential	$(1-\rho)+\rho\left(1-\omega\dfrac{\theta}{\mu-\lambda+\theta}\right)$

2.3. Revealed M/G/1 Queue

In this setting the arriving customer observes the length of the waiting line prior to the point at which his/her threshold value becomes fixed. The threshold setting process is modeled as a draw from a prespecified distribution, the parameters of which depend on the length of the line that the customer observes upon arrival. The manager of such a service delivery system is concerned with the PSC over all possible states of the system, and the unconditional distribution F(x) is of primary importance. This distribution may be expressed as,

$$F(x) = \sum_{n=0}^{n=\infty} q_n F_n(x) \tag{5}$$

In other words, the PSC must be defined by calculating the PSC given that n people are in the system (F$_n$) as well as the probability that n people are present upon customer arrival (q$_n$.) We will deal with special cases applying Lemmas 1 and 2 which yield closed form results. (Other approaches to this problem for more general cases are discussed in the appendix.)

2.3.1. The Revealed M/G/1 Model When Y$_n$ ~Exp(θ). If the distribution of customer thresholds is exponential for all states of the system, i.e., Y$_n$ ~ Exp(θ_n) for all n, we can obtain closed form expressions for F$_n$(0) and F(0). Using Lemma 1, we conclude that

$$F_n(0) = \hat{H}_n(\theta_n), \tag{6}$$

where $\hat{H}_n$ is the L-S transform of H$_n$ and is given by

$$\hat{H}_n(s) = \hat{B}^{n-1}(s)\hat{R}(s) = \hat{B}^{n-1}(s)\,\frac{1-\hat{B}(s)}{m_1 s}. \tag{7}$$

From (6) and (7) we have

$$F_n(0) = \hat{B}^{(n-1)}(\theta_n)\,\frac{1-\hat{B}(\theta_n)}{m_1\theta_n}. \tag{8}$$

For this special case, it also holds that

$$F(0) = \sum_{n=0}^{\infty} q_n F_n(0) = \sum_{n=0}^{\infty} q_n \hat{H}_n(\theta_n)$$

$$= \sum_{n=0}^{\infty} q_n \hat{B}^{(n-1)}(\theta_n)\,\frac{1-\hat{B}(\theta_n)}{m_1\theta_n}. \tag{9}$$

Table 4 summarizes F$_n$(0) values for a variety of service time distributions derived using (8). A variety of

Table 4 Percentage of Satisfied Customers in Revealed Queues When *n* Customers are in the System, Customer Thresholds Y_n ~ Exp(θ_n) and Service Times Follow Various Distributions

Distribution of service times	Percentages of satisfied customers ($F_n(0)$)
Exponential ($1/\mu$)	$\left(\frac{\mu}{\theta_n + \mu}\right)^n$
Gamma (α, β)	$\frac{1}{\alpha\beta\theta_n}\left[\left(\frac{\alpha\beta}{\alpha\beta + \theta_n}\right)^{-\alpha n}\left\{1 - \left(\frac{\alpha\beta}{\alpha\beta + \theta_n}\right)^{\alpha}\right\}\right]$
Deterministic	$\left(\frac{1}{\mu\theta_n}\right)^n\left(\frac{\theta_n\mu - 1}{\theta_n\mu}\right)$
Erlang (k)	$\frac{k\mu}{\theta_n}\left(\frac{k\mu}{\theta_n + k\mu}\right)^{k(n-1)}\left(1 - \left(\frac{k\mu}{\theta_n + k\mu}\right)^k\right)$
Hyperexponential	$\frac{1}{\theta_n \sum_{i=1}^{m} \frac{\omega_i}{\mu_i}}\left(\omega_i \sum_{i=1}^{m} \frac{\mu_i}{\theta_n + \mu_i}\right)^{n-1}\left(1 - \omega_i \sum_{i=1}^{m} \frac{\mu_i}{\theta_n + \mu_i}\right)$
Generalized exponential	$\frac{\mu}{\theta_n\omega}\left(\frac{\omega\mu}{\theta_n + \mu}\right)^{n-1}\left(1 - \frac{\omega\mu}{\theta_n + \mu}\right)$

sources may be used to estimate q_n values for M/G/1 queues such as the tabulated results included in many textbooks. Thus, we can calculate $F_n(0)$ directly for many settings making it easy to calculate F(0).

2.4. The Revealed M/M/1 Model for Various Distributions of Y_n

With exponential service times we have, $\hat{R}(s) = \hat{B}(s)$ and $\hat{H}_n(s) = \hat{B}_n(s)$. For the special case where Y_n~Exp(θ_n), we have

$$F_n(0) = \left(\frac{\mu}{\mu + \theta_n}\right)^n, \text{ and} \tag{10}$$

$$F(0) = \sum_{n=0}^{\infty} (1 - \rho)\rho^n F_n(0) = (1 - \rho) \sum_{n=0}^{\infty} \left(\frac{\rho\mu}{\mu + \theta_n}\right)^n. \tag{11}$$

For cases in which Y_n is not exponentially distributed, we may make use of the fact that the waiting time is the sum of n exponential variables and solve the following equation

$$P[(W - X) \le 0] = 1 - P[(X - W) \le 0] \tag{12}$$

$$= 1 - \int_0^{\infty} F_x(w)g(w)\partial w$$

where g(w) is the p.d.f. of an Erlang distribution with parameters k and μ. $F_x(w)$ is the c.d.f. of the threshold X, evaluated at w. If we set $\theta = k\mu$, we may state this quantity as

$$F_n(0) = 1 - \frac{\theta^n}{(n-1)!}(-1)^n \frac{\partial^n \hat{F}(\theta)}{\partial\theta^n}. \tag{13}$$

In other words, if we know the nth derivative of the L-S transform of the threshold distribution, we can calculate $F_n(0)$. We then combine these values with q_n values to calculate F(0).

2.5. M/M/1 Queues With Exponentially Distributed Thresholds, Balking, and Reneging Customers

Our analysis may be extended to include both balking and reneging behavior among customers. For the special case in which we have an M/M/1 system where the probability of a customer balking can be stated as a function of the length of the line (1 - b_n), we know that the percentage of customers who balk is simply, $\beta = \Sigma_{n=1}^{\infty} q_n(1 - b_n)$. We label the rate at which customers renege when there are k customers in front of them as $\delta_k = \Sigma_{j=1}^{k} \delta_j$. If we label the percentage of arriving customers that do not balk or renege and are satisfied as $F'(0)$, we can now state that $F'(0) = \Sigma_{n=0}^{\infty} q_n b_n(\mu + \delta_j/\mu + 2\delta_j)^n F_n(0)$. (See appendix for additional details.) We label the percentage of reneging customers as Φ and state the value along with the PSC as

$$\Phi = \sum_{n=1}^{\infty} b_n q_n\{1 - \hat{W}_n(\delta)\}, \text{ and} \tag{14}$$

$$F(0) = F'(0)(1 - \beta - \Phi). \tag{15}$$

3. Observations on Hidden and Revealed M/G/1 Queues

For the case of hidden queues, we assume X ~ Exp(θ). For the case of revealed queues we assume that Y_n~Exp(θ/n^{α}), where $0 \le \alpha \le 1$. We focus upon this functional form because it appears reasonable to assume that Y_n should be both increasing and concave in n. When $\alpha = 0$ threshold levels are independent of n. We assume that $\alpha \le 1$, since it seems unreasonable to assume that customer thresholds would grow faster than n. Considering the scenarios discussed in Tables 2 and 4 we derive several useful results including the following.

Proposition 1. *For hidden queues when X ~ Exp(θ), or revealed queues when Y_n ~Exp(θ/n^{α}) the percentage of satisfied customers F(0) is:*

(i) a non-increasing, convex function of θ for the M/G/1 model;

(ii) a non-increasing, concave function of ρ for the M/G/1 model;

(iii) a non-decreasing, concave function of k for the M/E_k/1 model; and

(iv) a non-decreasing function of (1 - ω) for the M/GE/1 model.

Proposition 1 implies that when $Y_n \sim \text{Exp}(\theta/n^{\alpha})$, a shift in the mean of the distribution from which customers' expectations are drawn has a relatively sizable impact on PSC. This is particularly important when θ is low (part i) or when ρ is high (part ii). By increasing k from 1 (the exponential case) toward infinity (the deterministic case), we can display the impact that reducing service time variability has on the PSC. For example, consider the setting in which $\alpha = 0$, $\mu = 100$, θ is extremely small (0.00001), and ρ is extremely high ($\rho = 99.9\%$). Reducing service time variability has the greatest impact in the most extreme cases, such as this one. For this instance, changing k from 1 to infinity increases PSC by 16.9% (from 41.7% to 58.6%). Alternatively, an equivalent increase in PSC could have been obtained by dropping ρ to 99.5%. In this sense, we see that changing θ or ρ has a far greater impact on PSC than altering k. Figure 1 shows PSC values as a function of k when ρ is fixed at 0.7 for various levels of α. We note that changing α may produce a significant increase in PSC for any value of k and that the impact of increasing k is positive but marginal by comparison.

To understand the effects of reducing the variability of the distribution of customer thresholds, we focus upon the case in which expectations are drawn from an Erlang distribution with parameter k. Consideration of our expression for PSC leads to the following result.

Proposition 2. *For the M/M/1 model, the percentage of satisfied customers* $F(0)$ *is:*

(i) a non-decreasing, concave function of k, when X (or Y_n*)* $\sim E_k$

(ii) a decreasing, linear function of ω_i *when* $X \sim H_m(\omega_i, \theta)$

(iii) an increasing, linear function of $(1 - \omega_p)$*, when* $X \sim GE(\omega_{p,}\ \theta)$.

This result is graphically depicted in Figures 2 and 3. From Figure 2, we observe that the effect of a reduction in the variability of customer thresholds may be realistically labeled as marginal. For this example, a reduction from $C_x^2 = 1.0$ to 0.1 leads to an increase in the percentage of satisfied customers of 0.018. Figure 3 depicts results noted in parts ii and iii. Part ii relates to a setting where some portion of the customer base (ω_I) would best be described as impatient, meaning that any wait produces dissatisfaction. Part iii represents a setting in which the satisfaction of a fraction of customers ($1 - \omega_p$) is independent of the waiting time. The other customers have threshold values drawn from an exponential distribution with mean θ. In this case, the PSC rises linearly with ($1 - \omega_p$), when ρ is held constant. A unique line exists for each ρ value on Figure 3, and its slope rises as ρ rises. These two sets of curves are shown together to emphasize the finding that when all else is equal, growth in the "patient" segment of the population is more significant than an equal growth in the "impatient" segment of the population.

Figure 4 presents the result of an analysis of an M/M/1 system where customers balk upon finding a system containing n customers with probability ($1 - b_n$) and renege at a rate δ. The results shown here compare the PSC values under several scenarios. Case 1 assumes no balking or reneging. In case 2, we con-

Figure 1 F (0) Vs. k for M/Er(k)/1 System With $\rho = 0.70$ and $\theta = \mu = 100$ at Various Levels of α.

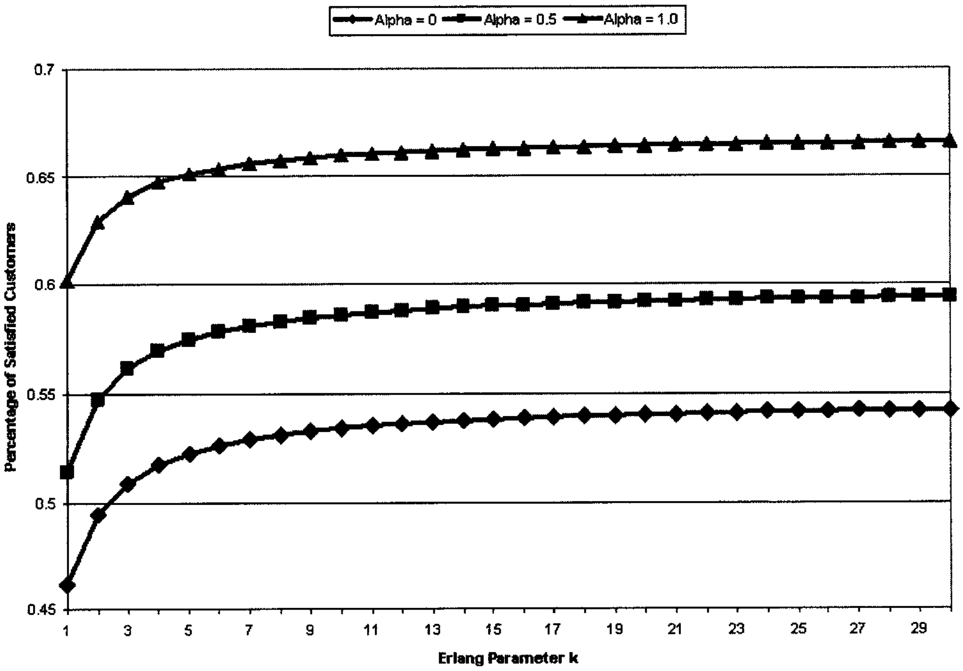
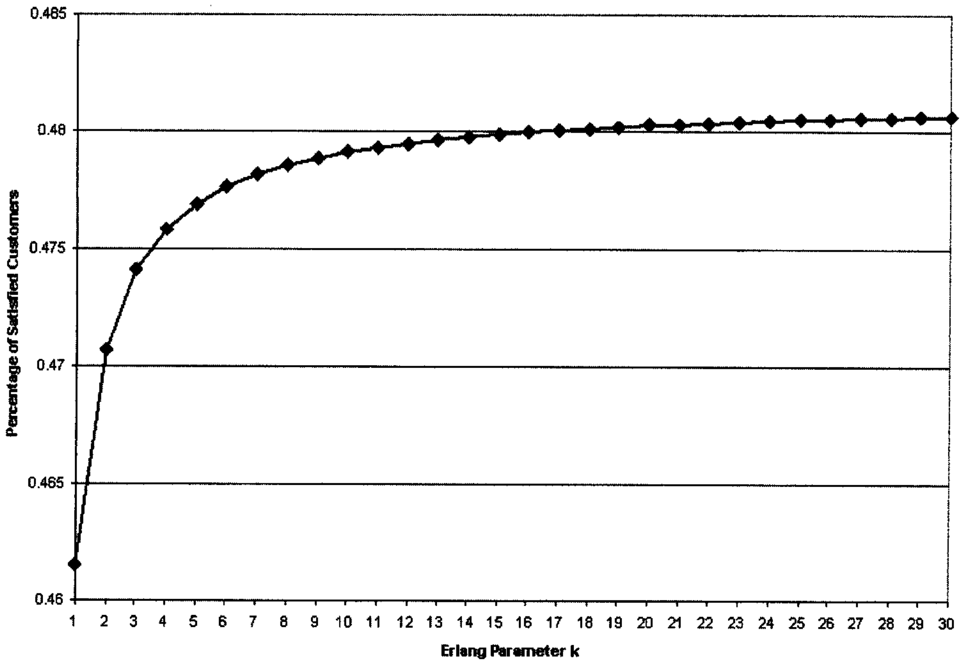


Figure 2 F (0) Vs. k for M/M/1 System With X ~ Er(k), $\theta = \mu = 100$, and $\rho = 0.70$.



sider balking at a rate of 5%. This value was suggested by research on call centers (Salzman and Merotha 2001) which found that roughly 5% of customers balk upon being put on hold even if the average waiting times are very short. In case 3, we consider reneging without balking, where δ =10, and in case 4, we treat both forms of customer loss. The results shown in Figure 4 indicate that balking and reneging serve to increase PSC values relative to the base case as arrival rates rise. Balking reduces the effective arrival rate, and reneging makes the line move faster for the customers that chose to endure the wait.

3.1. Hidden vs. Revealed Queues

Consideration of our metric for M/M/1 queues leads to the following result.

Proposition 3. *For revealed M/M/1 systems with $Y_n \sim Exp(\theta/n^{\alpha})$, and $0 \leq \alpha \leq 1$ the percentage of satisfied customers $F_n(0)$ is a monotone decreasing function of n.*

It is intuitive to expect that longer lines will produce longer waits, reducing customer satisfaction. Proposition 3 adds to this insight by showing that if customer thresholds grow no faster than the length of the wait-

Figure 3 F (0) for M/M/1 System With X ~ H or X ~ GE and $\theta = \mu = 100$ at Various Levels of ρ.

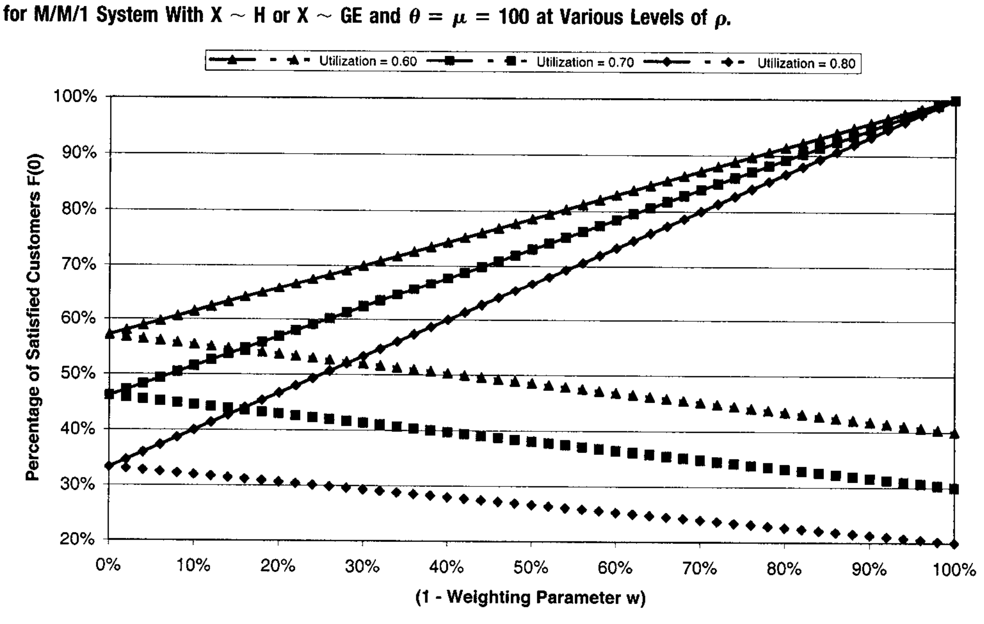
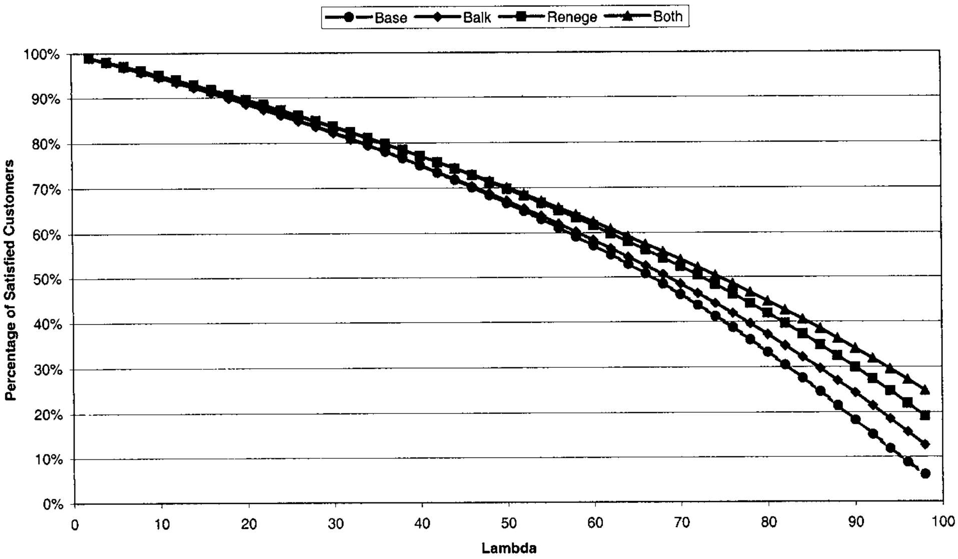


Figure 4 F (0) Vs. λ for M/M/1 System With $\theta = \mu = 100$, Balking, and Reneging.



ing line, then satisfaction rates decrease as the line length grows. Further analysis of $F_n(0)$ also leads to the following result.

Proposition 4. *For revealed M/M/1 systems with $Y_n \sim Exp(\theta/n^{\alpha})$, $0 \leq \alpha \leq 1$ the percentage of satisfied customers F(0) is a monotone increasing function of α and a decreasing function of ρ.*

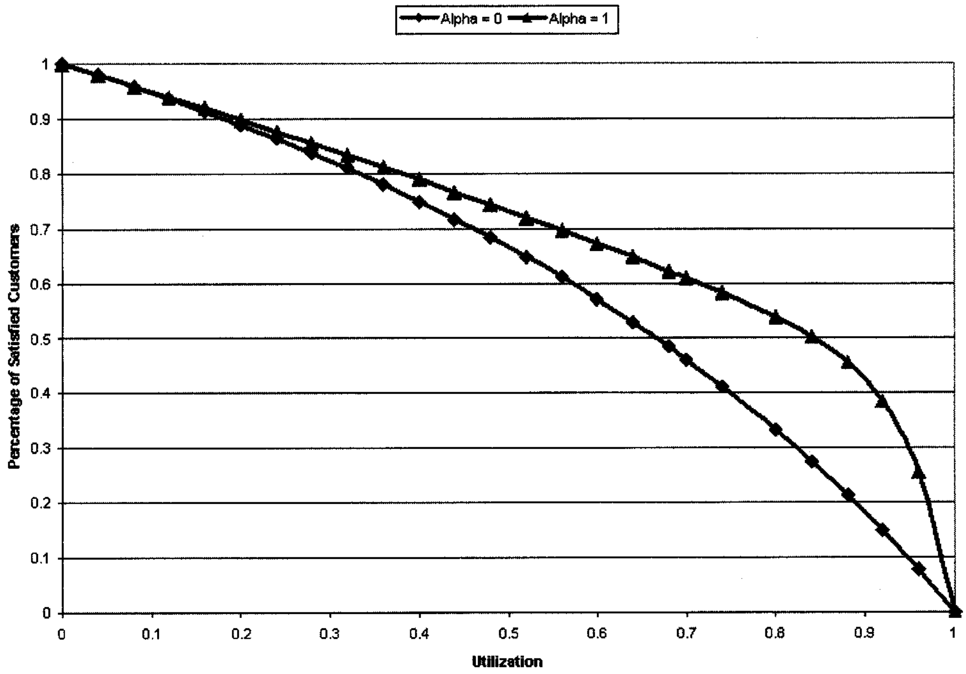
Figure 5 illustrates this result. It plots F(0) values as a function of utilization rates for the M/M/1 queue when $\alpha = 0$ or 1. It is apparent from the figure that the gains from using a revealed queue grow with utilization up to a point where ρ values are very high. (In this case, $\rho > 0.95$.)

4. Hidden and Revealed M/M/c Queues

4.1. Hidden M/M/c Queue

Consider a single line M | M system with c servers. When customer thresholds are drawn out of an exponential distribution with parameter θ, we can express the PSC as follows:

Figure 5 F (0) Vs. ρ for M/M/1 System With $\theta = \mu = 100$ at Various Levels of α.



$$F(0) = q_0 \sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + q_0 \frac{(\rho c)^c}{(c-1)!} \frac{\mu}{c\mu + \theta - \lambda}, \text{ where} \tag{16}$$

$$q_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \left(\frac{(c\rho)^c}{c!} \right) \left(\frac{1}{1-\rho} \right) \right]^{-1}. \tag{17}$$

(Additional details are available in the Appendix.)

4.2. Revealed M/M/c Queue

When customer thresholds are drawn out of an exponential distribution with its parameter θ_n, where n is the length of the waiting line upon the customer arrival, the conditional waiting time distribution can be expressed as,

$$H_n(x) = P\{W \le x | L = n\}$$
$$= \begin{cases} \dfrac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} & \text{if } n \ge c \\ U(x) & \text{if } n < c \end{cases} \tag{18}$$

where U(x) is the step function defined in Table 1. Using Lemma 1, we can easily calculate

$$F_n(0) = \begin{cases} \left(\dfrac{c\mu}{c\mu + \theta_n} \right)^{n-c+1} & \text{if } n \ge c \\ 1 & \text{if } n < c \end{cases}. \tag{19}$$

Thus, when $\rho = (\lambda/c\mu)$, the percentage of satisfied customers is

$$F(0) = \sum_{n=0}^{\infty} q_n F_n(0) = \sum_{n=0}^{c-1} q_n + \sum_{n=c}^{\infty} q_n \left(\frac{c\mu}{c\mu + \theta_n} \right)^{n-c+1}, \text{ and}$$

$$F(0) = q_0 \sum_{n=0}^{c-1} \frac{(\rho c)^n}{n!} + q_0 \sum_{n=c}^{\infty} \frac{\rho^n c^c}{c!} \left(\frac{c\mu}{c\mu + \theta_n} \right)^{n-c+1}. \tag{20}$$

4.3. Revealed M/M/c Queue With Balking and Reneging Customers

Analysis of the multi-server system with balking and reneging customers is surprisingly similar to that of a single server system because the transform of the waiting time distribution has a very similar structure (see Whitt 1999a). Specifically, we may state the L-S transform of the waiting time as,

$$F_n(0) = \hat{W}_n(s) = \prod_{j=0}^{k} \left(\frac{c\mu + \delta_j}{c\mu + \delta_j + s} \right) \tag{21}$$

In these equations, k indexes the position in the line and c is the number of servers; therefore, n = c + k. We also note that the q values are modified to reflect a multi-server system as follows,

$$q_n = q_0 \lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{n\mu} \text{ for } 0 < n < c, q_n$$
$$= q_0 \lambda^n \prod_{i=1}^{c-1} \frac{b_{i-1}}{n\mu} \prod_{c}^{\infty} \frac{b_{i-1}}{c\mu + \delta}, \text{ for } c \le n < \infty, \text{ and } q_0$$
$$= \left(1 + \sum_{n=1}^{c-1} \lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{n\mu} + \sum_{n=c}^{\infty} \lambda^n \prod_{i=1}^{c-1} \frac{b_{i-1}}{n\mu} \prod_{c}^{\infty} \frac{b_{i-1}}{c\mu + \delta} \right)^{-1}. \tag{22}$$

4.4. Observations on Hidden and Revealed M/M/c Queues

For both hidden and revealed multi-server systems, the following result may be stated.

Proposition 5. *For both the hidden and revealed M/M/c models, the percentage of satisfied customers, F(0) is:*

(i) a non-increasing, concave function of ρ (assuming λ or μ is fixed);

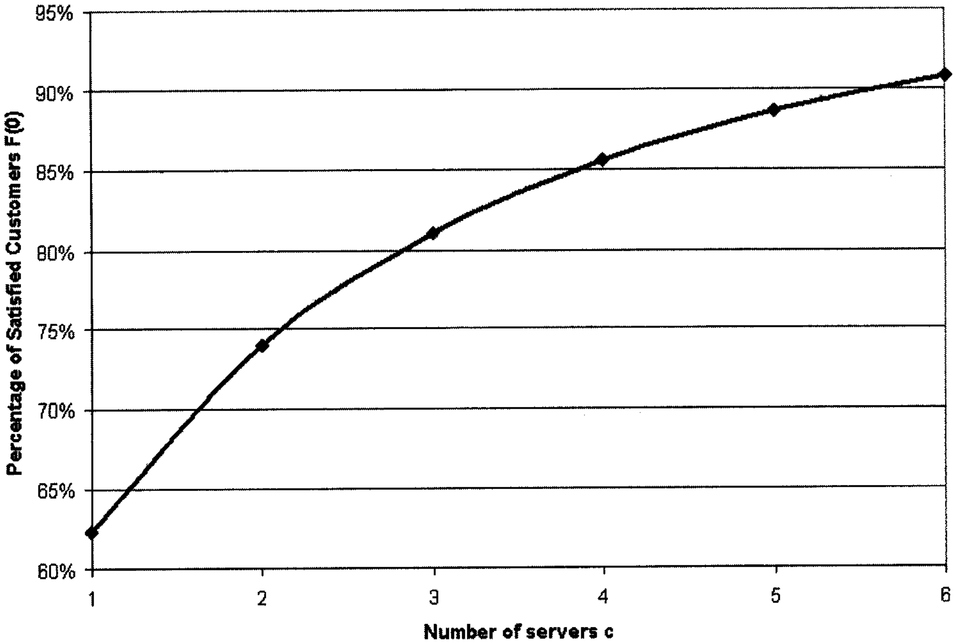
(ii) a non-increasing, convex function of θ.

Figures 6 and 7 are illustrative of this result for hidden M/M/c queues. Figure 6 shows that for a fixed total system capacity (i.e. cμ = a constant), the PSC increases as the number of servers increases. This strongly implies that designing a system with more, but slower servers leads to higher levels of customer satisfaction. Figure 7 shows that the PSC tends to become less sensitive to a shifting of the mean of the customers' threshold distribution as the number of slow servers is increased. This added benefit further advocates the use of a multi-"slow"-server system over a single "fast" server queue.

In Table 5, we report the percentage of satisfied customers for a hidden M/M/c model with X ~ Exp (θ) and for a revealed M/M/c model with Y_n ~ Exp (θ/n).

The table reports results for different levels of utilization and for different numbers of servers c (cμ is constant for all cases). From the results shown, we can infer that revealed M/M/c queues tend to dominate M/M/1 queues in terms of the PSC. This result is entirely reasonable given the observation that while the distribution of the number of customers in the system is stochastically smaller in the single-server case, the reverse holds for the number of customers in the queue (see Wolff 1989, p. 258). In other words, visitors to the revealed single server queue are likely to see longer lines, and the PSC decreases as a revealed queue gets longer.

Figure 6 F (0) Vs. θ for M/M/c System With ρ = 0.70 and Various Levels of c.



A result analogous to Proposition 4 can be stated for revealed M/M/c systems with $Y_n \sim$ Exp (θ/n).

Proposition 7. *For revealed M/M/c systems with $Y_n \sim Exp(\theta/n)$, the percentage of satisfied customers, $F_n(0)$ is a monotone decreasing function of n.*

The proof is omitted for brevity, but follows exactly the same steps as the proof of Proposition 4.

5. Conclusions and Managerial Insights

Our analysis and stated propositions lead to several insights that should prove useful for the management of waiting line systems.

Insight 1. *Reducing the psychological cost of waiting promises great payoffs by increasing the percentage of satisfied customers.*

Figure 7 F (0) Vs. c for M/M/c System at ρ = 0.70.

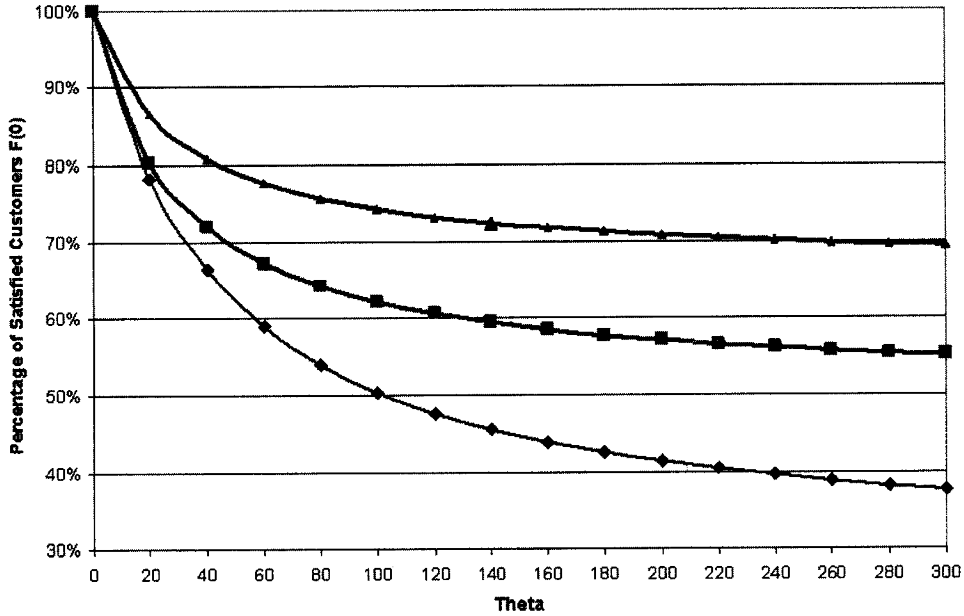


Table 5 Percentage of Satisfied Customers of Hidden and Revealed M/M/c Models for Various Levels of System Utilization for Various Values of c

Hidden M/M/c with $X \sim \text{Exp}(\theta)$

	Utilization ρ				
Number of Servers c	0.20	0.50	0.80	0.90	0.98
1	0.8889	0.6667	0.3333	0.1818	0.0392
3	0.9863	0.8421	0.4607	0.2572	0.0564
6	0.9997	0.9752	0.7646	0.5374	0.1561

Revealed M/M/c with $Y_n \sim \text{Exp}(\theta/n)$

	Utilization ρ				
Number of Servers c	0.20	0.50	0.80	0.90	0.98
1	0.8976	0.7320	0.5090	0.3474	0.0943
3	0.9932	0.9223	0.7095	0.5161	0.1531
6	0.9997	0.9787	0.8064	0.5821	0.1661

* For all cases $c\mu = 100$ and $\theta = c\mu$.

** The reported numbers are the percentage of satisfied customers $F(0)$.

If the service provider can alter customers' thresholds, then the firm is likely to produce a significant increase in the number of satisfied customers. Our results also show that increasing the portion of the population who don't mind waiting significantly increases values of F(0). The distribution of threshold values may be affected by a variety of means including providing information to reduce uncertainty, entertaining customers, making them more comfortable, or creating a greater sense of value for the end product.

Insight 2. *For many systems, reductions in the variability of service times or in the variability of customer thresholds have a lower than anticipated impact on the PSC.*

This is not an argument that reduction in service time variability is unimportant. However, our results do imply that reducing variability is less significant than intuition might suggest. Extremely impatient customers are not likely to be satisfied with any noticeable wait, and extremely patient customers are easy to satisfy. Our analysis suggests that the greatest impact arises from repositioning the threshold value of the "average" customer.

Insight 3. *An analytical estimation metric for PSC is critical for the management of many service delivery systems.*

As mentioned in prior research, including Kleinfeld (1988) managers routinely hear requests and advice to spend more money and offer more service. Our approach presents an objective and measurable basis for decisions regarding such policies. It is a simple matter to review survey data before and after a change is made to access its impact, but it is a far different matter to offer rationally developed predictions about satisfaction levels prior to a commitment to make such changes. Our metric provides one approach to help evaluate alternatives.

Insight 4. *Revealing information on waiting line lengths may be beneficial for service systems if it alters customer thresholds and lines are not likely to become very long.*

The impact of revealing queue length may be significant if customers alter their thresholds in a manner favorable to the firm. For cases in which ρ is low, such shifts are of little consequence. On the other hand, if ρ is moderately high; say between 60 and 85%, then positive α values can provide a meaningful increase in F(0). For systems with very high ρ values; say over 85%, the impact of customers raising their threshold values in response to viewing the length of the line is overwhelmed by the fact that the waits are very long. Another impact of providing information about the length of the queue is that some customers who would have been dissatisfied with the wait may chose to balk instead, perhaps to return during off-peak times. If this occurs, then the total percentage of customers who are satisfied actually rises.

Insight 5. *Assuming the same total system capacity, multiple slow server configurations have higher proportions of satisfied customers than do configurations with a single fast server.*

Significant improvements in customer satisfaction occur when systems move from a single "fast" server to two or three "slower" servers even if no decrease in system utilization results. This occurs because a one-to-one correspondence between line length and the means of the distribution of threshold values results in a scenario in which waits grow faster than customer patience when they view and wait in longer lines. The result is that even though pooling results in less waiting, it can also result in a lower PSC because customers are waiting in longer lines.

In summary, our approach emphasizes the fact that the objective of increasing customer satisfaction can profitably be approached by altering the service system to "provide more service" or by altering the customers' attitude toward the waiting experience. We have developed and presented a number of relatively simple analytic models that managers can use to evaluate the impact of policies which change the service delivery system or customer thresholds. These ideas may prove quite useful in the design and management of service delivery systems.

Appendix

Proof of Lemma 1.

$$P\{I - J \le 0\} = \int_0^\infty G_1(y)\theta e^{-\theta y} dy = \theta \frac{\gamma(\theta)}{\theta} = \gamma(\theta).$$

F(0) for Hidden Queues:

For a stationary M/G/1 system, the inversion formula of the Pollaczek-Khinchine (P-K) waiting time transform is known. (See Kleinrock 1975, p. 201, equation 5.111.)

$$h(t) = \sum_{n=0}^{\infty} (1-\rho)\rho^n r^{(n)}(t) \qquad \text{(A1)}$$

where r(t) is the stationary residual service time density given by

$$r(t) = \frac{d}{dt} R(t) \text{ and } R(t) = \frac{1}{m_1}(1 - B(t)) \qquad \text{(A2)}$$

and $r^{(n)}$ represents the n-fold convolution of r with itself.

We can evaluate h(t) numerically by applying the Laguerre transform to equations (A1) and (A2). For a detailed reference on the use of Laguerre transformation, see Keilson and Nunn (1979). When the distribution function of X is Gamma, hyperexponential, folded normal, their convolution, their mixture, or combination, it is convenient to evaluate the bilateral Laguerre transform of W – X numerically to describe the distribution of the random variable Z. For a detailed reference describing this approach, see Keilson, Nunn, and Sumita (1981).

F(0) for Revealed Queues

Recall that Y_n is the threshold value given that n people are in the line when the new customer arrives. We must consider the following distribution function: $F_n(x) = P\{W - Y_n \le 0 | L = n\}$. In particular, we are interested in calculating $F_n(0)$, which represents the probability that a customer arriving in the system when there are n customers ahead of him/her will be satisfied, i.e., $W \le Y_n$. Since the randomness of Y_n arises from the heterogeneity of the customer population, we assume that Y_n and W are conditionally independent given L. Therefore, we fix $Y_0 = 0$.

Suppose that a customer observes n customers ahead of him/her at the time of arrival. Because the "Poisson arrival sees time averages" (Wolff 1982), the residual service time of the customer being served, if any, at a customer arrival epoch has the distribution function R(t) given in (2). Thus, the distribution function of the waiting time given that an arriving customer sees n customers in the system is

$$H_n(x) \equiv P\{W \le x | L = n\} = \begin{cases} B^{(n-1)} * R(x) & n \ge 1 \\ U(x) & n = 0 \end{cases}, \qquad \text{(A3)}$$

where $B^{(n)}$ is n-fold convolution of B with itself, * represents a convolution operation, and U(x) is the step function previously defined.

When the distribution function of Y_n is Gamma, hyperexponential, folded normal, their convolution, their mixture, or combination of those, the distribution function $H_n(x)$ can be numerically evaluated using the bilateral Laguerre transform (see Keilson, Nunn, and Sumita 1981).

To complete the calculation of (5), we also need the stationary queue length distribution $(q_n)_{n=0}^{\infty}$ of our M/G/1 queue. Its calculation is based on the classical Pollaczek-Khinchine queue length formula (see Kleinrock 1975):

$$Q(z) \equiv \sum_{n=0}^{\infty} q_n z^n = \hat{B}(\lambda - \lambda z) \frac{(1-\rho)(1-z)}{\hat{B}(\lambda - \lambda z) - z}, \qquad \text{(A4)}$$

where $\hat{B}(s)$ is the L-S transform of B(x). Let us assume that the service time distribution function has a rational L-S transform. It is known that this class of distribution functions is "dense" (Kingman 1966) in the sense that every distribution function can be well approximated by a distribution function with rational L-S transform. Therefore, our assumption does not impose any serious restrictions in practice. In this case, $\hat{B}(\lambda - \lambda z) = (N(z)/D(z))$. We note that N(z) and D(z) are both polynomials. From (A4) we obtain,

$$Q(z) = \frac{(1-\rho)N(z)(1-z)}{N(z) - zD(z)}. \qquad \text{(A5)}$$

To invert (A3), we apply a straightforward power series expansion. Let

$$q_0 = Q(0) \text{ and } \hat{Q}_0(z) = Q(z). \qquad \text{(A6)}$$

Given $\hat{Q}_n(z)$ and q_n, let

$$\left.\begin{aligned} \hat{Q}_{n+1}(z) &= \frac{Q_n(z) - q_n}{z} \\ q_{n+1} &= \hat{Q}_{n+1}(0) \end{aligned}\right\}. \tag{A7}$$

Then, the recursive formula (A7), with the initial condition (A6) gives the series expression of (A4). From the series expression and (5), we can evaluate F(x).

F(0) with Balking and Reneging for M/M/1 Queues

The analysis of both hidden and revealed queues for such settings is very similar, thus their treatment is presented together here. Balking results in an effective arrival rates that take into account customers who abandon the queue before joining for any significant length of time. We label this rate $\lambda_n = \lambda b_n$ where $0 \le b_n \le b_0 = 1$. When the length of the queue is hidden, we can assume that b_i is a constant b for all i > 0 in hidden queues. For M/M/1 systems, we know (see Gross and Harris 1998, p. 94) that,

$$q_n = q_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu} = q_0 \left(\frac{\lambda}{\mu}\right)^n \prod_{i=1}^{n} b_{i-1}. \tag{A8}$$

For cases with reneging customers, we follow the approach in Gross and Harris (1998), p. 95, to devise q_n values where reneging rates (r(n)) are functions of n. Combining the effects of balking and reneging on q_n we have,

$$q_n = q_0 \lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{\mu + r(n)} \ (n \ge 1), \text{ and } q_0$$

$$= \left(1 + \sum_{n=1}^{\infty} \left\{\lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{\mu + r(n)}\right\}\right)^{-1}. \tag{A9}$$

We calculate the percentage of customers that balk as

$$\beta = \sum_{n=1}^{\infty} q_n(1 - b_n). \tag{A10}$$

When customers renege at a constant rate, r(n) = δ, this is equivalent to having customers leave the system after exponentially distributed amounts of time with rate δ. When a customer finds n = c + k customers in the system, the delay can be represented as the sum of k + 1 independent exponential random variables. Note that these variables are not identically distributed unless $\delta = \mu$. Reneging occurs at some rate δ_k for each position in the queue and other customers leave the system after being served at some rate μ. The total reneging rate when there are k = n – c customers in the queue is $\delta_k = \Sigma_{j=1}^{k} \delta_j$. Thus, the waiting time distribution is the sum of k + 1 independent exponential random variables with different means. The L-S transform of this distribution can be written as

$$\hat{W}_{k+1}(s) = \prod_{j=0}^{k+1} \left(\frac{\mu + \delta_j}{\mu + \delta_j + s}\right). \tag{A11}$$

For hidden systems when customers renege at a constant rate, δ_j is independent of j, so we may drop the subscript. For the case of a single server, application of Lemma 1 implies,

$$F_n(0) = \hat{W}_n(\theta) = \left(\frac{\mu + \delta}{\mu + \delta + \theta}\right)^n. \tag{A12}$$

$\hat{W}_n(\delta)$ given by (A11) states the percentage of customers who enter a system containing n customers and do not balk, nor renege. If we label the percentage of arriving customers that do not balk or renege and are satisfied as $F'(0)$, we can now state that,

$$F'(0) = \sum_{n=0}^{\infty} q_n b_n \left(\frac{\mu + \delta_j}{\mu + 2\delta_j}\right)^n F_n(0). \tag{A13}$$

We assume that customers who balk or renege are not satisfied. We label the percentage of reneging customers Φ. We can now state Φ and PSC = $F(0)$ as,

$$\Phi = \sum_{n=1}^{\infty} b_n q_n \{1 - \hat{W}_n(\delta)\}, \text{ and} \tag{A14}$$

$$F(0) = F'(0)(1 - \beta - \Phi). \tag{A15}$$

For additional details, see Rao (1968).

F(0) for Hidden M/M/c Queue

We know that $F(0) = \Sigma_{n=0}^{c-1} q_n + \Sigma_{n=c}^{\infty} q_n(c\mu/(c\mu+\theta))^{n-c+1}$. Expansion and rearranging of terms yields, $F(0) = \Sigma_{n=0}^{c-1} q_0(\lambda^n/n!\mu^n) + \Sigma_{n=c}^{\infty} q_0(\lambda^n c^c/\mu^n c! c^n)(c\mu/(c\mu+\theta))^{n-c+1} = q_0\ \Sigma_{n=0}^{c-1}\ ((\lambda/\mu)^n)\ (1/n!) + q_0\ \Sigma_{n=c}^{\infty}\ (\rho^n c^c/c!)(c\mu/(c\mu + \theta))^{n-c+1}$, where, $\rho = (\lambda/c\mu)$. Therefore, $F(0) = q_0\ \Sigma_{n=0}^{c-1}\ ((c\rho)^n/n!) + q_0\ ((\rho c)^c/c!)\ (c\mu/(c\mu+\theta))\ \Sigma_{n=0}^{\infty}\ (\rho\ (c\mu/(c\mu+\theta))^n$. Simplification leads to $F(0) = q_0\ \Sigma_{n=0}^{c-1}\ ((c\rho)^n/n!) + q_0\ ((\rho c)^c/c!)(c\mu/(c\mu+\theta))(c\mu+\theta)/(c\mu+\theta-\lambda)$. Rearranging terms leads to the stated form.

Proof of Proposition 4. If this is true for $\alpha = 1$, it is also true for all $0 < \alpha < 1$. Setting $\alpha = 1$, and $\theta_n = \theta/n$, then according to (10), we have

$$F_n(0) = \left(\frac{n\mu}{n\mu + \theta}\right)^n.$$

Then

$$\ell nF_n(0) = n(\ell nn\mu - \ell n(n\mu + \theta)) \rightarrow \frac{d\ell nF_n(0)}{dn}$$

$$= \ell nn\mu - \ell n(n\mu + \theta)n\left(\frac{\mu}{n\mu} - \frac{\mu}{(n\mu + \theta)}\right) \rightarrow \frac{\frac{dF_n(0)}{dn}}{F_n(0)}$$

$$= \ell nn\mu - \ell n(n\mu + \theta) + \frac{\theta}{(n\mu + \theta)}.$$

Let $E_n = \ell nn\mu - \ell n(n\mu + \theta) + (\theta/(n\mu + \theta))$. Then $\lim_{n\rightarrow 0} E_n = -\infty$ and $\lim_{n\rightarrow 0} E_n = 0$. Since

$$\frac{dE_n}{dn} = \frac{m}{n\mu} - \frac{\mu}{(n\mu + \theta)} - \frac{\mu\theta}{(n\mu + \theta)^2}$$

$$= \frac{\mu\theta^2}{n\mu(n\mu + \theta)^2} > 0,$$

E_n is a monotone increasing function and $E_n < 0$ for all $n > 0$. Thus,

$$\frac{dF_n(0)}{dn} < 0 \text{ for all } n > 0.$$

References

Boulding, W., A. Kalra, R. Staelin, V. A. Zeithaml. 1993. A dynamic process model of service quality: From expectation to behavioral intentions. *Journal of Marketing Research* **30** 7–27.

Carmon, Z., J. G. Shanthikumar, T. F. Carmon. 1995. A psychological perspective on service segmentation: The significance of accounting for consumers' perceptions of waiting and service. *Management Science* **41**(11) 1806–1815.

Dube-Rioux, L. B. H. Schmidt, R. LeClerc. 1988. Consumer's reactions to waiting: When delays affect the perception of service quality in *Advances in Consumer Research,* Srull, E. (ed.). Association for Consumer Research, Provo, Utah, pp. 59–63.

Green, L., P. Kolesar. 1987. On the validity and utility of queueing models of human service systems. *Annals of Operations Research* **9** 469–479.

Gross, D., C. M. Harris. 1998. *Fundamentals of queueing theory,* 3rd edition. Wiley Inter-Science, New York, New York.

Hall, R. W. 1991. *Queueing methods for services and manufacturing.* Prentice Hall, Englewood Cliffs, New Jersey.

Inman, R. R. 1999. Empirical evaluation of exponential and independent assumptions in queueing models of manufacturing systems. *Production and Operations Management* **(4)** 409–432.

Keilson, J., W. R. Nunn. 1979. Laguerre transform as a tool for the numerical solution of integral equations of convolution type. *Applied Mathematics and Computation* **5** 313–359.

Keilson, J., W. R. Nunn, U. Sumita, 1981. Bilateral laguerre transform. *Applied Mathematics and Computation* **8** 137–174.

Kingman, J. F. 1966. *On the algebra of Queues.* Methuen's Suppl. Review Series on Applied Probability.

Kleinfeld, N. R. 1988. Conquering those killer queues. *New York Times*, September 25, pp 1–11.

Kleinrock, L. 1975. *Queueing systems (Volume I: Theory)*. John Wiley, New York, New York.

Maister, D. 1985. The psychology of waiting lines in *The Service Encounter,* John Czepiel, Michael Solomon, Carol Suprenant (eds.). Lexington Books, Lexington, Massachusetts 113,23.

Rao, S. S. 1968. Queueing with balking and reneging in M G 1 systems. *Metrika* **12** 173–188.

Wardell, D. Goodale, J., Gupta, J. N. D. 2002. Empirical distributions of interarrival times and service times in call centers. Unpublished working paper.

Whitt, W. 1983. Deciding which queue to join: Some counterexamples. *Operations Research* **34** 55–62.

Whitt, W. 1999a. Improving service by informing customers about anticipated delays. *Management Science* **45**(2) 192–207.

Whitt, W. 1999b. Predicting queueing delays. *Management Science* **45**(6) 870–888.

Wolff, R. W. 1989. *Stochastic modeling and the theory of queues.* Prentice Hall, Englewood Cliffs, New Jersey.

Wolff, R. W. 1982. Poisson arrivals see time averages. *Operations Research* **30** 223–231.