Fluid Approximation of a Priority Call Center With Time-Varying Arrivals

Ahmad D. Ridley, Ph.D.
William Massey, Ph.D.
Michael Fu, Ph.D.

In this paper, we model a call center as a preemptive-resume priority queue with time-varying arrival rates and two priority classes of customers. The low priority customers have a dynamic priority where they become high priority if their waiting-time in queue exceeds a given service-level time. The performance of the call center is measured by the mean number in system for the two customer classes. A fluid approximation is proposed to estimate the mean number in system for each class. The quality of the approximation is tested by comparing it with a stochastic simulation model of the system. Finally, using the fluid approximations, we discuss how to compute the mean number in system for each class and estimate the overall staffing level, or number of agents.

Introduction

Call centers have become the primary channel of customer interactions, sales, and service for many businesses. Traditional call center performance modeling is based on simple Markovian queueing models, developed to analyze telephone traffic across the Public Switched Telephone Network (PSTN). Closed-form solutions for most of these queueing models are only available for steady-state behavior. Thus, these solutions are not applicable to practical call centers because of the time-varying, or transient, behavior of the arrival call process. In addition, these traditional models become problematic as call centers progress from handling only voice calls to handling multiple types of "calls," such as voice, e-mails, faxes, and Web chat sessions. In other words, they do not accurately analyze the performance of modern, multimedia call centers.

To better measure the performance of multimedia call centers over time, we developed mathematical fluid approximations instead of using simple Markovian queueing models. We modeled a multimedia call center as a preemptive-resume priority queue with time-varying arrival rates and two priority classes of customers. The high priority customer class consists of regular telephone, or voice, calls, while the low priority customer class contains e-mail "calls." The low priority calls have a dynamic priority where they are upgraded to a high priority customer based on their service level. Usually, this service level is the probability that the waiting-time in queue is less than a given time duration.

The call center performance measured by our fluid approximation is the mean number of calls in the system for each customer class. Our preemptive-resume, time-varying model is a complex one that cannot be solved with traditional Markovian queueing techniques. The fluid approximations are computed using an asymptotic scheme where the ratio of the offered load to the number of servers remains constant. The mean number in system for both customer classes is a solution to a system of differential equations. In this paper, we investigate the effectiveness of the fluid approximations through a comparison with the stochastic, discrete-event simulation method and measure the difference between the mean number in system computed using both methods. We also discuss our results and describe our future efforts for computing the mean virtual delay for both customer classes.

Call Center Overview

Traditionally, customers contacted a call center by talking to a customer service representative (CSR), or agent, over the telephone. Now, customers can contact an agent over the Internet, either by e-mail or chat session. Many companies use call centers—banks, financial institutions, information technology (IT) help desks, and government agencies. The managers of these call centers attempt to provide their customers with efficient and convenient service. However, their job is much more difficult today because there are far more products and services being sold and supported than a few years ago. Thus, the managers struggle to deliver different service levels to different types of customers with different needs and issues.

The advancement in call center technologies not only provides more benefits, but also more challenges.

For example, current technologies provide managers greater flexibility in routing and queueing calls by prioritizing certain types of incoming calls and allowing customers to access call agents with different skill sets. The manager's job of scheduling agents and satisfying multiple customer service levels therefore becomes more complex.

Technical Components of a Call Center

A traditional call center has several main components; namely, an automatic call distributor (ACD), an interactive voice response (IVR) unit, desktop computers, and telephones. [1] The ACD is a telephone switch located at a customer's premises and provides methods for the distribution of customer calls. [2] There is a finite number of trunks (i.e., telephone lines) connecting the ACD to the PSTN.

As customer calls arrive, the ACD receives and routes them either to the IVR unit where customer transactions are handled automatically, or to an idle CSR, who provides the necessary service. If no CSR is available, the calls are placed in a queue (i.e., on hold). The CSR responds to the calls routed to them using their telephone and desktop computer. For example, if the agent is answering a telephone call, that agent can access the customer information database through the desktop computer. The heart of a traditional call center is this dynamic routing of a new or pending call by the ACD to the most appropriate and available CSR. This call routing or assignment process must take into consideration such factors as the call priority, call arrival time, and CSR skills and availability. It requires dynamic, real-time management of all CSR skill levels and availability, the call/caller identity and status, and customer information databases. Therefore, the flow of an arriving call through a call center can be complex.

Call Center Modeling

The basic structure of the call center can be described as a finite capacity, multi-server system. Customer calls arrive at the call center at varying rates on a finite number of trunks. These calls are terminated at the ACD switch and are routed to a group of agents. In a multimedia call center, these calls can be voice, e-mail, fax, or (eventually) video.

Queueing Methods

We use queueing models to analyze the performance of the call center. Current analytical models applied in practice are based on traditional Markovian queueing models. A Markovian model is represented symbolically as M/M/N/L, where

- M = the arrival process as a stationary Poisson process, where the inter-arrival times of customers, or calls, are exponentially distributed with a mean constant call rate (note that M_t identifies a non-stationary Poisson process, where the arrival call rates vary over time)
- M = the service times of the calls as exponentially distributed random variables
- N = the number of servers, or call agents, at the queue
- L = the number of spaces available in the system, i.e., the total number of servers and queue spaces; in call center terminology, this value L is known as the total number of trunk lines available to calls

Problem Setting

We are studying a complex call center system for which simple Markovian queueing models do not apply. Our goal is to develop alternative methods to estimate the transient performance for our system, rather than approximating them with steady-state M/M/N/L queueing systems. Specifically, we developed a fluid model and a separate simulation model to approximate the mean number in system for both high and low priority customers at different points in time. Our call center is a help desk with two-customer classes and a preemptive-resume priority queue discipline. The high priority customer class consists of voice calls, while the low priority customer class consists of e-mails. Here, we assumed that there are enough telephone lines to prevent any call blocking. Also, we assumed that the service level for the high priority class is high enough that no calls abandon the system. Note that in a general call center environment these assumptions are not always valid.

In our model, the customers are served from two distinct virtual queues. The customers from the lowest priority class, i.e., the e-mails, will leave the low priority queue and enter the higher priority queue based on a specified service level parameter. In this regard, the low priority calls will have dynamic priorities. Our goal was to show that the fluid approximations of the call center performance are close to the actual performance, as measured by a discrete-event simulation model. Figure 1 displays the queue diagram of our model.

We define the notations in Figure 1 as follows:

- $\lambda_i(t)$ represents the arrival rate for class i customers into queue i, (i=1,2)
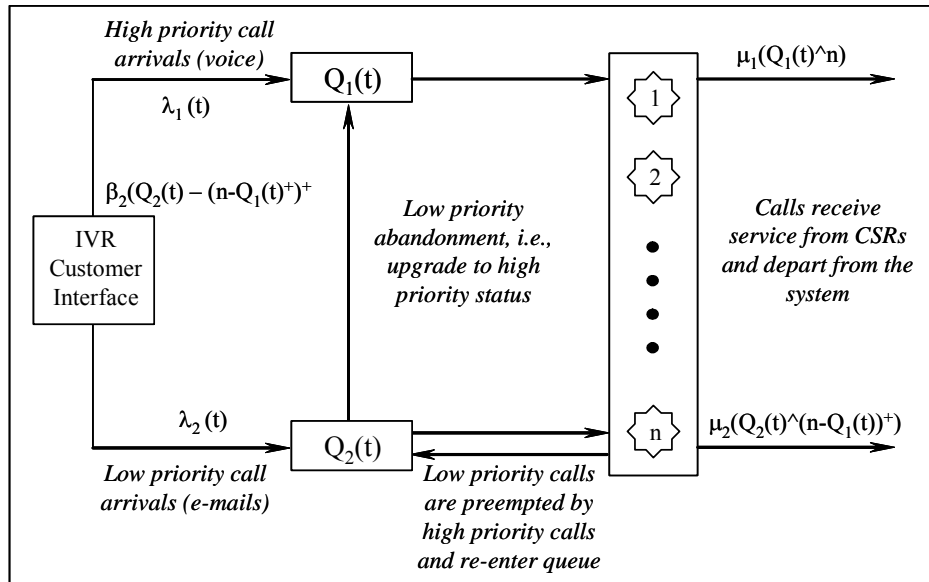- $Q_i(t)$ represents the number of class i customers in the system

Figure 1. Two-Class, Preemptive-Resume Priority Queue With Low Priority Abandonment

- β_2 is the abandonment rate of low priority customers out of the low priority queue
- n is the number of servers, or CSRs, in the system, which remains constant over time
- $x \wedge y$ represents the minimum between x and y
- $(x-y)+$ represents the maximum between 0 and x-y

Markovian Models Based on Erlang

Through his research on the telephone network in the early 1900s, Erlang showed that the arrival process of calls over the network to any destination could be modeled as a Poisson process. [3] Although these models are primarily used for producing daily call forecasts and agent work schedules, they do attempt to explain the randomness that exists in call centers. This randomness is caused by the variability of call arrival patterns and call durations. The most common queueing models use call volumes, call handling times, and the number of agents to compute the average waiting times for customers during steady-state. The Erlang-B and Erlang-C models are two traditional Markovian models used in practice to estimate the performance of a call center. In both, the arrival process of calls to a call center is modeled as a Poisson process. The Erlang-B model can be represented as the M/M/n/n queue. Thus, the inter-arrival call times are exponentially distributed with mean $1 / \lambda$, and the call service times are also exponentially distributed with mean $1/\mu$. There are n servers, or call agents with a first-come first-serve queue discipline, and a system capacity of n calls. Here, $\rho = \lambda / (\mu*n)$, where the quantity λ/μ is defined as the offered load of the traf-

fic. Because this model assumes a finite system capacity of n calls, a call may be blocked from entering the call center. This blocking probability, β_n, is an important performance measure and is given by the following steady-state formula: [3]

$$\beta_n = P(\{\text{all n servers are busy}\}) = ((\lambda / \mu)^n / n!) / (\Sigma_k \ \lambda / \mu)^k / k!), \ k = 0,\ldots,n. \quad (1)$$

The above formula is also referred to as the Erlang-B, or Erlang Loss formula.   The Erlang-C model can be represented as the M/M/n queue. The model is useful when $\lambda/\mu < n$, where λ is the mean arrival call rate, μ is the mean call service rate, and n is the number of agents. Because this model assumes an infinite system capacity of calls, there is no probability of blocking. Although calls may enter the call center when all n servers are busy, they must wait in queue before receiving service.   In this model, this probability of waiting in queue (i.e., probability of call delay), or $P(D>0)$, is important to measure and is given by the following formula: [3]

$$P(D > 0) = P(\text{at least n calls in system}) = ((n \ \rho^n / n!)(1/(1- \rho)) / ((\Sigma_k (n \ \rho^k / k!) + (n \ \rho^n / n!)(1 / (1-\rho))), \ k =0,\ldots,n-1, \quad (2)$$

where D is the delay of a customer call. Also, the mean delay, E[D], is given by: [3]

$$E[D] = (P(D>0)*(e^{-(n-\rho)\mu t})) / (\mu*(n-\rho)).$$

The above formulas for the MM/n and M/M/n/n queues, and those for the M/M/N/L queue, are used in

practice to estimate the number of agents required to satisfy customer service levels, and the average delay experienced by customers. These Markovian queueing models are often based on the following key simplifying assumptions.

- Every call is of the same type
- Every agent can handle calls equally fast
- The arrival rates do not vary over time, and the system enters steady-state under certain conditions
- Calls are queued on a first-come-first-serve basis

Unfortunately, under these assumptions, the M/M/N/L queueing models can sometimes differ significantly from the real-world call center performance measures.

Note that the Erlang-B model underestimates the blocking probability of the system by assuming no queue forms. Also, the Erlang-C model overestimates the average delay (in queue) experienced by calls by assuming that the queue size is infinite. Therefore, these models, and Markovian models in general, have some limitations in representing call center systems.

Fluid Approximations

Service systems models, such as call center models, belong to the class of stochastic service network models. These network models form a special family of non-stationary Markov processes where parameters such as arrival and service rates are time-dependent. More importantly, these models have functional strong laws of large numbers and functional central limit theorem results for the number of customers in the system and the waiting time in queue. [4] The results are developed using an asymptotic limiting process, where the number of servers is scaled up in response to a scaling up of the arrival rates; in other words, the number of servers and arrival rates are multiplied by the same factor.

These limit theorems lead to a tractable set of network fluid approximations in the form of a system of ordinary differential equations (ODEs). By numerically solving these differential equations using either the Euler or Runge-Kutta method, we can compute values for the mean number in system at specific points in time. More importantly, this technique allows us to approximate solutions of models that are otherwise analytically intractable using Markovian queueing techniques. [4] Therefore, an alternative, possibly more robust, method can be developed and applied to the performance analysis of service systems, such as call centers.

Two-Customer Class Model

Our fluid approximations for the mean number in the system will be derived for the two-customer class, preemptive-resume priority, $M_t/M/n$ queue. Since the high priority customers can preempt the lower priority ones, these customers will essentially receive service as if no other type of customer is present in the system. Thus, the high priority customer class results will be almost the same as the results for the single customer class. The only difference is the dynamic priority process for the low priority customers, where these customers can leave their queue and enter the high priority queue as a high priority customer. This process adds an extra term to the differential equations describing the process for the high priority customers.

Asymptotic Limit Theorems

The results and theorems presented in this section are adapted from those stated by Mandelbaum, Massey, and Reiman [4] and Mandelbaum, Massey, Reiman, et al. [5] However, customers are now grouped into two classes: high priority and low priority. High priority customers are labeled as class-1 customers, while low priority customers are labeled as class-2 customers. Thus, all of the random variables of the stochastic processes, discussed in Mandelbaum, Massey, Reiman, et al., [5] are now random vectors. For example, the $M_t/M/n$ number in system process $Q = \{Q(t) \mid t \geq 0\}$ must be defined for two-customer classes. The random variable $Q(t)$ is now defined as the random vector $Q(t)=\{Q_1(t), Q_2(t)\}$, for all positive real numbers t. [6] Here, the random variables, $Q_1(t)$ and $Q_2(t)$, are the corresponding quantities for class-1 and class-2 customers, respectively.

The limit theorem for the functional strong law of large numbers can be restated for our model, as follows.

Theorem 5.2

$$\lim_{\eta \to \infty} \{1/\eta\} Q^\eta = Q^{(0)}, \text{ (almost surely, i.e., with probability 1)}$$

where the convergence is uniform on compact sets of t and η is the scale factor for the arrival rate, $\lambda(t)$, and number of servers, n. Moreover, $Q^{(0)} = \{Q^{(0)}(t) \mid t \geq 0\} = \{Q_1^{(0)}(t), Q_2^{(0)}(t) \mid t \geq 0\}$ is uniquely determined by the initial function value $Q^{(0)}(0)$ and the differential equations:

$$\frac{dQ_1^{(0)}(t)}{dt} = \lambda_1(t) - \mu (Q_1^{(0)}(t) \wedge n) - \beta[Q_2^{(0)}(t) - (n - Q_1^{(0)}(t))^+]^+; \qquad (3)$$

$$\frac{dQ_2^{(0)}(t)}{dt} = \lambda_2(t) - \mu [Q_2^{(0)}(t) \wedge (n - Q_1^{(0)}(t))^+]^+ - \beta[Q_2^{(0)}(t) - (n - Q_1^{(0)}(t))^+]^+, \qquad (4)$$

where $[Q_2^{(0)}(t)-(n-Q_1^{(0)}(t))^+]^+$ is the number of customers in the low priority queue.

This theorem states rigorously that $Q^\eta \cong \eta Q^{(0)}$ for large η, where $Q^{(0)}$ is called the fluid approximation for Q^η. In other words, as the offered load and number of servers becomes large, the fluid approximation provides a good estimate to the mean number in system, $Q(t)$. The proof of the theorem is given in Kleinrock, 1975. [3]

Simulation Model

The final method used to compute the mean number in system for high and low priority customers is a discrete-event simulation model. Our model approximates an $M_t/M/n$ queue where the arrival process is a time-varying Poisson process, and the service times are exponentially distributed. Also, there are two classes of customers that arrive to the system, where the lower class is upgraded to the higher class status, as discussed earlier.

Discrete-Event Simulation

Discrete-event simulation deals with representing a time-varying system with a series of state variables that change instantaneously at distinct points in time. [6] In mathematical terms, the system can change at only a countable number of points in time. The state variables in our simulation are the number of class-1 and class-2 calls in the call center, or system. The events that change the state of the system are the arrival and departure of customers, or calls, into and out of the call center. Therefore, discrete-event simulation is used to implement the queueing model of our call center. Note that we implemented our simulation model using the C-programming language.

Arrival Process

One of the main components of the stochastic simulation is the arrival process. We chose to approximate the true arrival rate function as a piecewise linear function over a set of disjoint 30-minute time subintervals $[t_a, t_{a+1}]$ which partition the overall finite-time horizon interval $[0, T]$, where $a=1,2, \ldots, m-1$, and m represents the number of 30-minute subintervals.

Since our model supports two types of customers, $\lambda(t)$ is the overall arrival rate and is defined as $\lambda(t) = \lambda_1(t) + \lambda_2(t)$, where the arrival rates for the high priority customers, $\lambda_1(t)$, and the low priority customers, $\lambda_2(t)$, also vary with time. We randomly determined the call type of each customer upon their arrival. Here, based on Poisson thinning, a customer will have call type i with probability $\lambda_i(t) / \lambda(t)$.

Comparison Results of Two Models

Our goal was to compare two different estimates of the mean number in system for both customer classes. Thus, we compared our results from the fluid model to the simulation model for the $M_t/M/n$, two-class, preemptive-resume, dynamic priority queue.

Call Center Data

We began our computation of numerical results by defining the queueing model parameters. The parameter values were taken from a real-world, help desk call center, in which calls represent requests for IT support (e.g., network support, password resets, application support, etc.). The help desk was simulated over a 12-hour day in our fluid and simulation models. Thus, each independent replication simulates the performance of the help desk over the course of a day. All the rates used in the methods were per-minute rates. Note that in the fluid and simulation methods, a piecewise constant function was used for the time-varying arrival rate function, λ_t. The duration of each value of λ_t was 30 minutes. Thus, λ_t varies every 30 minutes during the 12 hours, or 360 minutes of our time horizon. Figure 2 contains a graph of our arrival rate function. Therefore, we derived our model parameters, such as arrival rates, service rates, service levels, and number of agents, from a real-world help desk call center.

Here, the high priority customer calls were telephone, or voice, calls, and the low priority customer calls were e-mails. Note that many call centers today handle both voice calls and e-mails. However, there are some challenges in gathering information about e-mail customer interaction with call centers. For example, managers collect more detailed information on parameters for telephone calls than for e-mails. In other words, some parameters, such as service rates, are not often collected for each e-mail that arrives to a call center. Most call center managers simply use a "best effort" approach to handle e-mail customers. Also, in some call centers, the group of agents that respond to e-mails is not the same as the group that respond to voice calls. Note that our model assumed that a single group of agents had the necessary skills to respond to both voice calls and e-mails.

For the multi-server, non-stationary queues, we set n=20, where n is the number of agents, or servers. Since we are using asymptotic limits for the fluid approximations, we must scale both the arrival rates and the number of agents towards infinity in order to compute accurate fluid estimates. We used a scale factor of 25. The server utilizations vary over time between 0.1302 and 1.245, where the maximum value occurs from 8:30 to 9:00 AM and the minimum value occurs
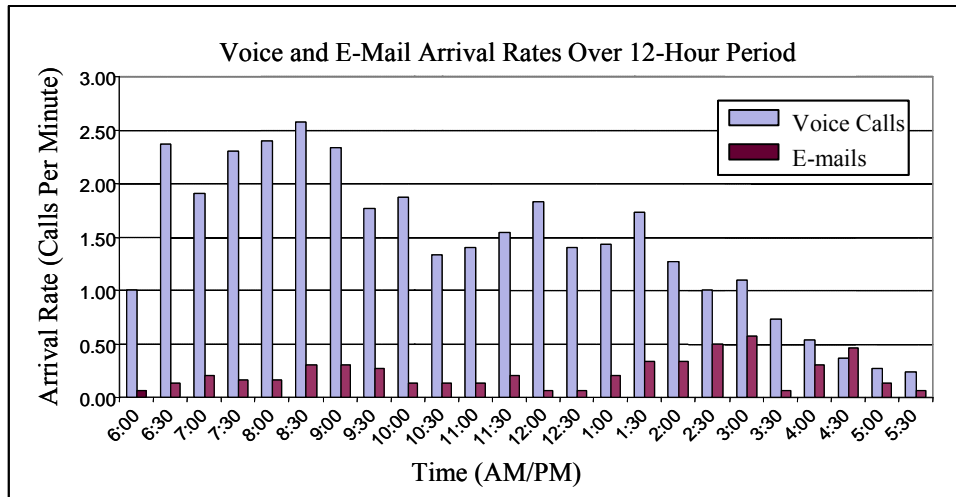
Figure 2. Arrival Rates for High Priority (Voice) and Low Priority (E-Mail) Call Classes

from 5:30 to 6:00 PM. Therefore, our system progressed through overloaded (>1, or unstable) and underloaded (<1, or stable) phases, as the arrival rates vary over time.

The mean service time for high priority calls was 8.69 minutes per customer, or equivalently 521.29 seconds per customer. The service rate for the high priority calls, μ_1, was the reciprocal of the mean service time, so $\mu_1 = 1 / 8.9 = 0.1151$ customers per minute. In our help desk, mean service times were not reported for the low priority, or e-mail, customers. Thus, we let the service rate for the low priority customers equal that of the high priority customers, which is not an unreasonable assumption. Therefore, we set $\mu_2 = \mu_1 = 0.1151$ customers per minute.

Numerical Results

We compared the numerical results between the fluid approximations and discrete-event simulation estimates for the mean number in the system for each customer class. The mean number in the system was estimated at several time points, t_i, where the t_i's are spaced 60 minutes apart over the time horizon. Figures 3 and 4 show the comparison of the mean number in the system results for the high and low priority customers between our fluid and simulation models. Note that for most of the t_i's, the fluid approximations are very close to the simulation estimates, for the high priority and low priority calls. In fact, as the offered load, which is a measure of the intensity of the calls to the help desk, varies over time, the accuracy of our fluid approximations remained good. For example, the largest loads occurred from 6:30 to 9:00 AM and the smallest loads occurred after 3:30 PM. However, our approximations were very good in all time periods, but did depend on the scale

factor. In other words, at a scale factor of 10, the fluid approximations began to converge to the simulation estimates. At a factor of 25, the fluid approximations were very close to the simulation estimates. Beyond a factor of 25, the approximations were not much better, suggesting that 25 was a good stopping point for our scale factor. Since the fluid approximations for the mean number-in-system were very close to their corresponding simulation values, we expected the estimates for the mean waiting-time in queue from both methods to be close as well. (Intuitively, the amount of time a customer waits in queue depends on the number of customers in the system, especially those ahead of the customer.)

Conclusions

We obtained fairly accurate fluid approximations to the simulation results for the mean number in system for the high and low priority customer classes. Note that the number of differential equations in our fluid approximations method is independent of the number of servers in the call center. Thus, the complexity of our fluid approximations method does not increase as the call center increases in size, or the number of agents increases. However, it is more likely that the simulation will increase in complexity as the call center becomes larger. Therefore, our fluid approximation is a much more scalable solution than the simulation.

Future Research

Currently, we use our fluid approximations to estimate the mean number in system for both customer classes. In our future research, we will use our fluid approximations and simulation model to determine the mean
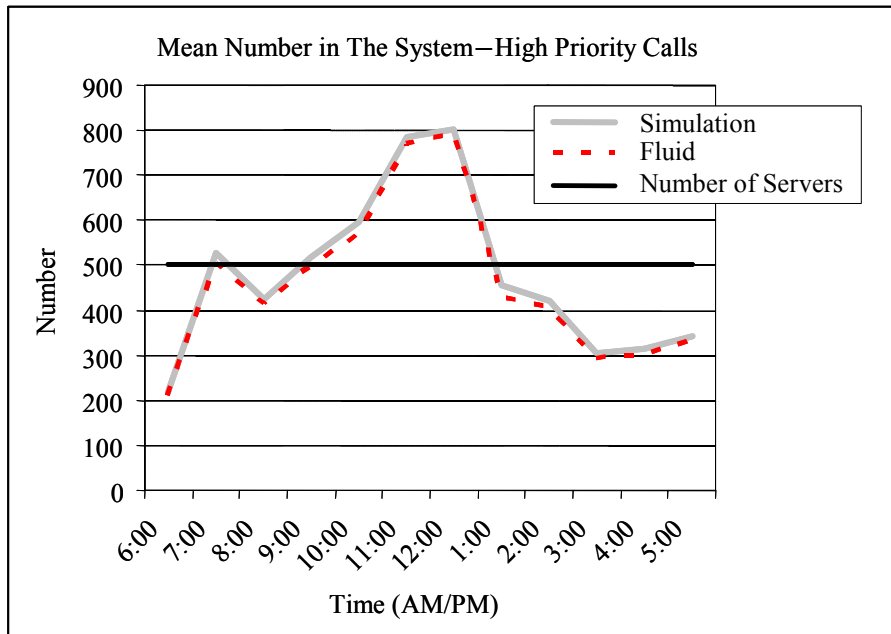
Figure 3. Fluid and Simulation Comparison for Mean Number in System—High Priority
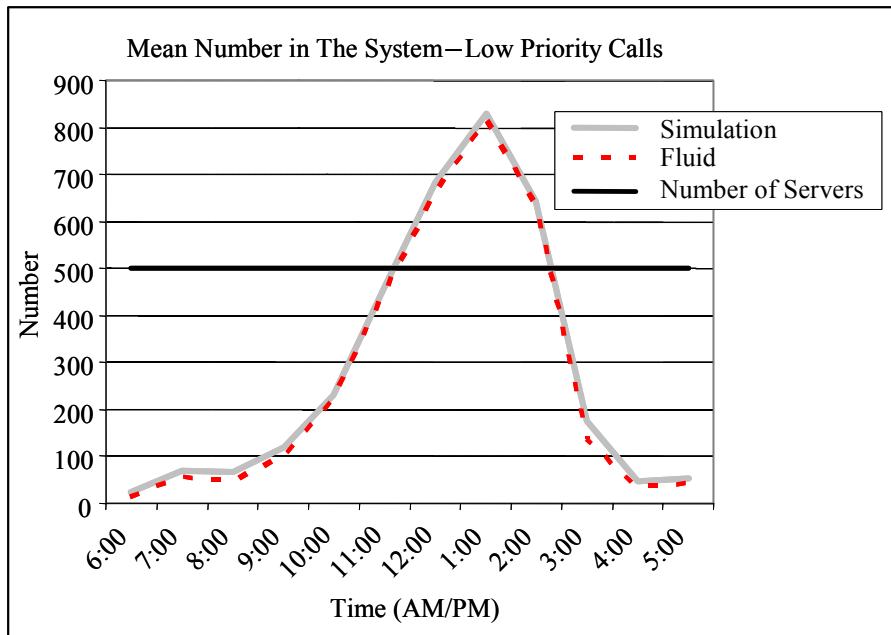

Figure 4. Fluid and Simulation Comparison for Mean Number in System—Low Priority

virtual waiting time of the high and low priority customers. Our mean waiting-time results will be an extension of the results in Mandelbaum, Massey, Reiman, et al. [5] from the single customer class case to the two-customer class one. The waiting-time computation for the low priority customers is more complex than the one for the high priority customers. If these low priority customers are preempted and move to the high priority queue, their waiting time will be a combination of their time in the low priority queue, their partial-service time before being prempted from

a server, and their waiting-time in the high priority queue.

We will then use the mean virtual waiting-time approximations to predict an actual staffing level to handle the overall number of customers. The criteria for changing the staffing level, or number of servers, in our model will be based on a comparison of the mean virtual waiting-time for each customer class to its corresponding target service level, or mean waiting-time. The simple staffing algorithm follows.

- Choose an initial staffing level, or value for the number of servers, and target service level for the high and low priority customers; these values are determined from our actual call center data.
- Compute the mean virtual waiting-time using the fluid approximations for each customer class.
- If the percentage of mean virtual waiting-times is greater than the target service level for either class, then increment the number of servers by 1.
- Repeat the second step until the target service level is satisfied for both classes of customers.
- We will use this predicted staffing level in our simulation. Finally, we will verify the accuracy of our staffing prediction by comparing the mean virtual waiting time from the simulation for each class with its corresponding target service level.

References and Notes

1. Bernett, H. G., M. J. Fischer, and D. M. B. Masi, "Internet Protocol/Public Switched Telephone Network Blended Call Center Performance Analysis," *The Telecommunications Review*, Volume 12, pp. 51–60, Mitretek Systems, 2001.
2. Hall, R. W., "Queueing Methods for Services and Manufacturing," Prentice Hall, 1991.
3. Kleinrock, L. "Queueing Systems, Volume I: Theory," John Wiley & Sons, 1975.
4. Mandelbaum, A., W. A. Massey, and M. I. Reiman, "Strong Approximations for Markovian Service Networks," *Queueing Systems*, Volume 30, pp. 149–201, 1998.
5. Mandelbaum, A., W. A. Massey, M. I. Reiman, R. Rider, and A. Stolyar, "Queue Lengths and Waiting Times for Multi-Server Queues with Abandonments and Retrials," *Proceedings of the Fifth INFORMS Telecommunications Conference*, 2001.
6. Law, A. and W. David Kelton, "Simulation Modeling and Analysis," McGraw-Hill, 2000.

About the Authors

Dr. Ahmad D. Ridley is a Senior Staff Engineer at Mitretek Systems. He has worked on several telecommunications projects under the General Services Administration's (GSA) Federal Technology Service (FTS) program, including Wire and Cable Services, Washington Interagency Telecommunications Services 2001, and FTS2001 contracts. His research interests include the application of applied mathematics to telecommunications network modeling and optimization. Dr. Ridley received his Ph.D. in Applied Mathematics at the University of Maryland, College Park; his M.S. degree in Applied Mathematics from the University of Maryland, College Park; and his B.A. degree in Mathematics from the University of Maryland, Baltimore County.
E-mail: aridley@mitretek.org

Dr. William Massey is "Edwin S. Wiley" Professor in the Department of Operations Research and Financial Engineering at Princeton University. His research interests include performance and pricing models for telecommunications systems, queueing systems with time-varying rates, and asymptotic analysis and stochastic bounds for stochastic networks. Dr. Massey is a member of many professional societies, to include the American Mathematical Society (AMS), Committee for African-American Researchers in the Mathematical Sciences (co-founder), Institute for Operations Research and Management Sciences (INFORMS), and the National Association of Mathematicians (lifetime member). He is a member of Phi Beta Kappa and Sigma Xi. Dr. Massey received his Ph.D. in Mathematics from Stanford University in 1981, and his A.B. degree in Mathematics, Magna Cum Laude, from Princeton University in 1977.
E-mail: wmassey@princeton.edu

Dr. Michael Fu is Professor of Management Science in the Decision and Information Technologies Department at the Clark School of Engineering. He has a joint appointment with the Institute for Systems Research and an affiliate appointment with the Department of Electrical and Computer Engineering, both in the Clark School of Engineering. His research interests include simulation modeling and analysis, production/inventory control, applied probability and queueing theory, with application to manufacturing and finance. He is co-author of the book, "Conditional Monte Carlo: Gradient Estimation and Optimization Applications," which was awarded the INFORMS College on Simulation Outstanding Publication Award in 1998. Dr. Fu received his Ph.D. in Applied Mathematics from Harvard University in 1989 and his M.S. degree in 1986. He was a National Science Foundation Graduate Fellow from 1995 to 1998; and he received his S.M. and S.B. degrees in Electrical Engineering, Computer Science, and Mathematics from MIT in 1985.
E-mail: mfu@rhsmith.umd.edu