Variational Optimization for Call Center Staffing

Robert C. Hampshire and William A. Massey
Department of Operations Research and Financial Engineering
Princeton University
rhampshi@princeton.edu and wmassey@princeton.edu

1. Introduction

According to Koole and Mandelbaum [5], almost 60 to 70 percent of the total costs for operating a call center involve wage and benefit expenses for personnel. It follows that determining the optimal amount of call center agents is of great interest to call center managers. This personnel or staffing problem has been studied by others (see [5] for extensive references). Previous work often assumes that an unlimited number of telephone lines are available to handle calls. This paper addresses both the staffing of agents and the provisioning of telephone lines by introducing a revenue and penalty structure. Our goal is to develop an approximate algorithm for designing a profit optimal staffing and provisioning schedule. Our method for determining the number of agents and telephone lines arises from variational optimization methods.

First, we model the call center as a multiserver queue with additional waiting spaces and abandonment. This queueing system is a special case of a natural class of queueing network models for call centers called *Markovian service networks*. They were identified and analyzed in Mandelbaum, Massey and Reiman [6]. Markovian service networks capture many important call center features such as time varying arrival rates, multiserver queues, service abandonments, as well as network routing due to service completions or service abandonments. Inspired by growing a business to match a corresponding growth in customer demand (as first used in Halfin and Whitt [3]), we can scale these stochastic queueing models so that they converge to a deterministic "fluid" model.

Now we add an economic structure to our queueing model for the call center. We assume that there is a reward for every successful service completion, a penalty for every abandoned call, and a cost for the number of agents and telephone lines used. We can then express the total profit for the call center as an integral functional of the time evolution for the number of customers in the system over a fixed time interval. We call this our *profit functional*.

In general, the fluid models for these Markovian service networks are *dynamical systems* whose evolution is governed by a set of non-linear, ordinary differential equations. The dimension of these equations correspond to the number of service nodes in the network. The queueing model for our call center is one-dimensional. The corresponding fluid model is governed by an ordinary differential equation. We then use variational calculus methods from the theory of optimal control to derive a optimal staffing and provisioning schedule from our analysis of the fluid approximation of the profit functional.

2. Markovian Service Network and Fluid Models

Our stochastic call center model, the *queueing system process* $Q = \{ Q(t) \mid t \geq 0 \}$, is defined by letting $\lambda(t) =$ non-homogeneous Poisson arrival rate at time t, $\mu =$ exponential service rate, $\beta =$ exponential "music"-abandonment rate, $\gamma =$ exponential "busy signal"-abandonment rate, $K(t) =$ number of additional telephone lines at time t, and $L(t) =$ number of call center agents at time t.

Now we construct the *uniformly accelerated* version of the model (see [6]) with scale factor $\eta > 0$. In the context of call centers, we can motivate this asymptotic scaling by considering the expansion of a business in response growing customer demand. The "size" of this call center business is given by the number of call center agents $L(t)$ and telephone lines $K(t) + L(t)$. Similarly the "size" of the aggregate customer demand is given by the arrival rate $\lambda(t)$. The service rate and abandonment rates μ, β and γ correspond to personal decisions made by individual customers that are independent of the total size of customer demand or the total size of the call center. This follows from the fact that the typical customer is unaware of both of these dimensions for the call center. Based on these assumptions, it it reasonable to scale the parameters λ, K and L upwards by η, but *not* to scale μ, β or γ. Now consider the limit of this process $Q^\eta \equiv \{ Q^\eta(t) \mid t \geq 0 \}$ as $\eta \to \infty$. By appealing to the strong law of large numbers, we can construct a deterministic approximation for the mean behavior of the unscaled system.

From the general theory for Markovian service networks (see [6]), it follows that whenever we have $\lim_{\eta \to \infty} Q^\eta(0)/\eta = Q(0)$, there is a deterministic process $Q^{(0)} \equiv \left\{ Q^{(0)}(t) \mid t \geq 0 \right\}$

such that

$$\lim_{\eta \to \infty} \frac{1}{\eta} Q^\eta = Q^{(0)} \quad \text{a.s.} \tag{2.1}$$

where this almost sure convergence is uniform over compact intervals of time. Moreover, this deterministic process $Q^{(0)}$ is a *dynamical system* that is governed by the differential equation

$$\frac{d}{dt} Q^{(0)}(t)$$
$$= \lambda(t) - \beta \cdot \left(\left(Q^{(0)}(t) - L(t) \right)^+ - \left(Q^{(0)}(t) - K(t) - L(t) \right)^+ \right)$$
$$- \gamma \cdot \left(Q^{(0)}(t) - K(t) - L(t) \right)^+ - \mu \cdot \left(Q^{(0)}(t) \wedge L(t) \right)$$

and referred to as a *fluid model*.

We now have a deterministic dynamical system that approximates our random queueing process. By adding on a pricing and cost structure, we can develop an approximate optimal profit design analysis by applying the variational calculus tools of optimal control theory to the fluid model.

3. Variational Analysis of the Fluid Profit Model

We add a pricing and cost structure to our call center queueing model with r = service completion reward per customer, s = music abandonment penalty per customer, τ = busy signal abandonment penalty per customer, $c(L)$ = total staffing cost rate for L agents, $d(K+L)$ = total provisioning cost rate for $K+L$ telephone lines. For the rest of the paper, we assume that K and L are non-negative functions that are *not* necessarily integer-valued. Using the fluid model $Q^{(0)}$ to approximate the mean queueing behavior for the stochastic call center model, our goal now is to find $Q^{(0)}$, K and L and an additional multiplier $x \equiv \{ x(t) \mid 0 \le t \le T \}$ so that we can maximize the integral

$$\mathcal{A}\left(Q^{(0)}, \dot{Q}^{(0)}, x, K, L \right) \equiv \int_0^T \mathcal{L}(Q^{(0)}, \dot{Q}^{(0)}, x, K, L)\, dt, \tag{3.1}$$

where

$$\mathcal{L}(Q^{(0)}, \dot{Q}^{(0)}, x, K, L)$$
$$\equiv r\mu \cdot \left(Q^{(0)} \wedge L \right) - s\,\beta \cdot \left(\left(Q^{(0)} - L \right)^+ - \left(Q^{(0)} - K - L \right)^+ \right)$$
$$- \tau\,\gamma \cdot \left(Q^{(0)} - K - L \right)^+ - c\,(L) - d\,(K+L)$$
$$+ x \cdot \left\{ \dot{Q}^{(0)} - \lambda + \beta \cdot \left(\left(Q^{(0)} - L \right)^+ - \left(Q^{(0)} - K - L \right)^+ \right) \right.$$
$$\left. + \gamma \cdot \left(Q^{(0)} - K - L \right)^+ + \mu \cdot \left(Q^{(0)} \wedge L \right) \right\}.$$

In the language of classical mechanics, our generalized profit integral $\mathcal{A}$ is the *action* and its integrand $\mathcal{L}$ is the *Lagrangian*. The latter has the units of *energy* for a physical system and plays the role of a *profit rate* for our call center model. The fluid approximation $Q^{(0)}$ for the total number of customers in the call center plays the classical mechanical role of the *position* variable. Using the *Hamiltonian* reformulation of the Lagrangian, the multiplier x is the *generalized momentum* variable. Whenever $Q^{(0)}$ and x are extremal solutions

for the action, then the *Euler-Lagrange equations* apply and give us

$$x = \frac{\partial \mathcal{L}}{\partial \dot{Q}^{(0)}} \quad \text{and} \quad \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{Q}^{(0)}} = \frac{\partial \mathcal{L}}{\partial Q^{(0)}} \quad \Longrightarrow \quad \dot{x} = \frac{\partial \mathcal{L}}{\partial Q^{(0)}}. \tag{3.2}$$

If x corresponds to momentum, then $\dot{x}$ corresponds to the classical mechanical term of *force*. For call centers, (3.2) tells us that $\dot{x}$ is the *marginal profit rate per customer density*. It follows that x is the *marginal profit per customer density*. It measures of the impact on the total profit of an additional customer joining the system.

Assume that c and d are both non-negative, increasing, concave functions, our optimization problem reduces to the analysis of three simpler, "competing" Lagrangians $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$. We say that one of these Lagrangians is *dominant* at time t, if it is the largest of the three.

THEOREM 3.1. *Given $Q^{(0)}$ and x, construct K^* and L^* as follows:*

$$K^*(t) = \begin{cases} Q^{(0)}(t) & \text{if $\mathcal{L}_2$ is dominant,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

and

$$L^*(t) = \begin{cases} Q^{(0)}(t) & \text{if $\mathcal{L}_3$ is dominant,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

It holds that if $Q^{(0)}$, x, K^* and L^* maximize $\int_0^T \mathcal{L}\, dt$, then having $\mathcal{L}_1$, $\mathcal{L}_2$ or $\mathcal{L}_3$ being dominant corresponds respectively to $Q^{(0)}$ and x solving the Euler-Lagrange equations

$$\dot{Q}^{(0)} = \lambda - \gamma Q^{(0)} \quad and \quad \dot{x} = (x - \tau)\,\gamma, \tag{3.5}$$

$$\dot{Q}^{(0)} = \lambda - \beta Q^{(0)} \quad and \quad \dot{x} = (x - s)\,\beta - d'(Q^{(0)}). \tag{3.6}$$

or

$$\dot{Q}^{(0)} = \lambda - \mu Q^{(0)} \quad and \quad \dot{x} = (x + r)\,\mu - c'(Q^{(0)}) - d'(Q^{(0)}). \tag{3.7}$$

One immediate consequence of this theorem is that $K^*(t)$ and $L^*(t)$ are *complementary variables*, i.e. $K^*(t) \cdot L^*(t) = 0$.

To test our optimal schedule, we simulate the Markovian model of the call center. The profits obtained from this Markovian model by simulation are then compared to the optimal fluid profits. To provide evidence that this schedule is asymptotically optimal, we construct alternate schedules that are random perturbations of the optimal schedule. They are created randomly and are applied to the simulated Markovian model.

Figure 1 shows the staffing schedule and cost for this numerical case. Note that the staffing schedule goes through the three Lagrangian modes. Both sets of graphs show that the average profit for the simulated queue with the optimal profit schedule is greater than the average profits from the perturbed schedules.
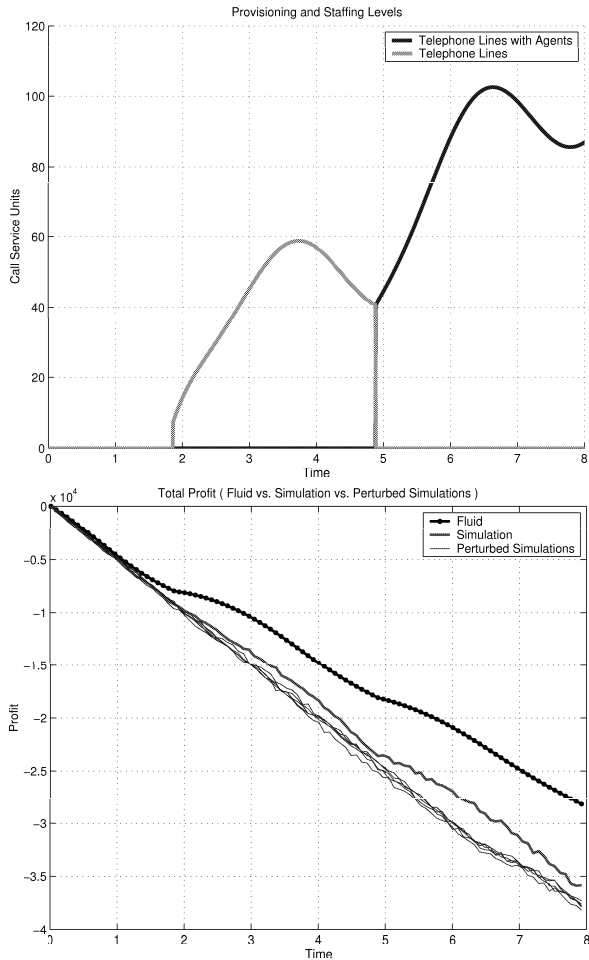
Figure 1: **Non-Profit example with** $\lambda(t) = 100.0 + 30.0 \cdot$
$\sin 0.75\pi t$, $\mu = 1.0$, $\beta = 2.0$, $\gamma = 10.0$, $r = 0.0$, $s = 38.9543$,
$\tau =$**39.0023**, $c(L) = 39.9 \cdot L$ **and** $d(K + L) = 0.1 \cdot (K + L)$.

4. REFERENCES

[1] Borst S., Mandelbaum A., and Reiman M. I.
"Dimensioning Large Call Centers," unpublished,
2001, Downloadable from
http://iew3.technion.ac.il/serveng.

[2] Gans, N., Koole, G. and Mandelbaum, A., "Telephone
call centers: Tutorial, review and research prospects."
Manufacturing and Service Operations Management,
Volume 5, pp. 79-141, 2003.

[3] Halfin, S. and Whitt, W. "Heavy-Traffic Limits for
Queues with Many Exponential Servers," Operations
Research, 29, pp. 567–587, 1981.

[4] Harris, C. M., Hoffman, K. L. & Saunders P. B.
"Modeling the IRS Telephone Taxpayer Information
System," *Operations Research*, Volume 35, Number 4,
pp. 504–523, July-August 1987.

[5] Koole, G. and Mandelbaum, A., "Queueing Models of
Call Centers: An Introduction," unpublished, Oct
2001, Downloadable from
http://www.cs.vu.nl/obp/callcenters or
http://iew3.technion.ac.il/serveng.

[6] Mandelbaum, A., Massey, W.A. and Reiman, M.I.
"Strong approximations for Markovian Service
Networks." *Queueing Systems*, 30 (1998) M.I. pp.
149-201.