Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

SIMULATION OF A CLAIMS CALL CENTER: A SUCCESS AND A FAILURE

Roger Klungle

AAA Michigan
Business Operations Analysis
1 Auto Club Drive
Dearborn, MI 48126, U.S.A.

ABSTRACT

This paper addresses the call center management process and the role of simulation in this process. Strengths and weaknesses of workforce management systems and the Erlang-C model are reviewed and the role of discrete event simulation is highlighted. An application in an insurance claims call center is utilized to show the effectiveness of simulation in evaluating call center designs and also the difficulties in selling selected results to management. The paper concludes with some lessons learned about the call center process, discrete event simulation, workforce management systems, and the specific claims application.

1 INTRODUCTION

Recent technology advances and the faster pace of change in business environments have made call center management a rapidly growing industry. Flexibility in call center design and processing is now the rule rather than the exception. The days of the "one size fits all" call center is fast eroding. With mass customization becoming more common place, calls can now be easily prioritized and routed to specific agents with multiple skill sets.

To deal with this myriad of issues now faced by call center managers, many organizations use workforce management systems as a key analysis tool. These systems forecast volumes, determine staffing levels and schedules, and monitor adherence to these schedules. However there are some real difficulties with these systems due to their static nature. In contrast, discrete event simulation is a more flexible and dynamic tool that can be of significant value in developing strategies and analyzing alternatives. A more detailed examination of these difficulties can be found in Klungle (1998).

2 APPLICATION BACKGROUND

A membership organization that has been in business for 80 years, AAA Michigan has over 1.6 million members in its auto club and insures over 1.2 million automobiles. Besides competing in the insurance market in Michigan, AAA also provides Club, Travel and Financial services. Club services include auto touring and emergency road service, Travel provides cruise packages, airline travel, and other vacation packages, and Financial issues credit cards.

In addition to an extensive branch office network throughout the state, seven call centers operate to provide the following services to members:

- Emergency Road Service (1 center) - Dispatches towing contractors to stranded motorists.
- Member Services (1 center) - Provides insurance service and sales, group insurance service, cellular phone sales, and other general assistance.
- Claim Service Centers (4 centers) - Takes initial reports on all claims taken by phone and handles all smaller claims to completion.
- Central Travel (1 center) - Takes orders for Triptiks, Tourbooks, and Maps by phone and provides airline ticket sales service.

Within this environment, simulation has been a key analysis tool for process re-engineering and continuous improvement, and for providing critical information for management decisions (Klungle and Maluchnik 1997). Although the focus of this paper is on call center applications, discrete event simulation is used for projects throughout the organization from the design of branch offices to the determination of road service equipment and location requirements.

3 CALL CENTER MANAGEMENT PROCESS

3.1 Steps

The effective management of a call center operation generally involves four major steps as shown in Figure 1 above. In this process, "Trunk Space Requirements" refers to the determination of the number of phone lines required to receive the volume of incoming calls and the number of queue slots needed to hold the calls during peak time periods.

Of the four steps, the business forecast is the most critical. It is the basis upon which trunk space and staff calculations will be made which are used to schedule the workforce to meet service levels and other performance criteria. More details on these forecasting issues can be found in Klungle and Maluchnik (1998).

3.2 Performance Measures

A key performance measure is the Service Level, which is defined as the percentage of time (% of half-hour periods) that a specified service goal is met. The 80/20 rule is generally used as the service goal where 80% of the calls are to be answered within 20 seconds. Thus a 90% Service Level means that the 80/20 rule is met for at least 90% of the half-hour periods. This service level goal is critical to the determination of appropriate staffing levels when using Erlang-C. One of the problems with using the service level criteria as defined is that during slow periods of a day (or season) where only a minimum staff is required (e.g. late afternoons and midnights), the 80/20 goal is almost always met. These large number of half-hour periods at low volumes are then weighted equally with the fewer half-hour periods where volumes are high and service levels are low, giving a distorted picture of the true quality of service. Other key performance measures include:

- Average speed of answer (ASA)
- Agent utilization
- Abandonment rate (reneging)
- Average length of call
- Percent answered without waiting

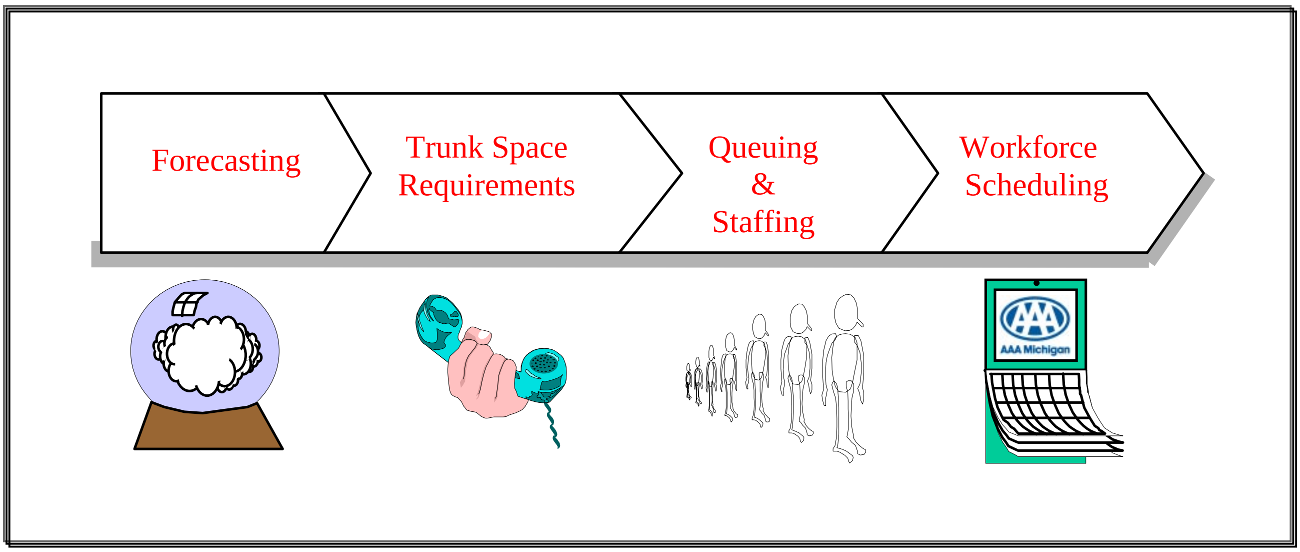

Figure 1: The Call Center Management Process

4 WORK FORCE MANAGEMENT SYSTEMS

4.1 Forecasting

In order to effectively run a call center, forecasts are generally needed at several levels. Long-term forecasts, such as yearly and monthly predictions, are used for budgeting and staff planning, planning operational changes, training, and scheduling vacations. Mid-term forecasts such as weekly and daily are needed for workforce staffing and scheduling. Short-term forecasts, usually every hour or every half hour, predict how well a call center is staffed for the current day.

Most workforce management systems use exponential smoothing to include trend and seasonal components (Winter's Model). This approach is fine if the time series is relatively stable which is truer for longer term forecasting (monthly or yearly). However, in many call centers such as Claims and Emergency Road Service (ERS), these volumes are highly volatile in the short term, resulting in large forecast errors for operational purposes. These large errors eventually lead to considerable over and under staffing and poor customer service. In such situations causal relationships (regression analysis) are more appropriate if the causal variables can be identified. These differences in approaches have a significant bearing on how forecasts are handled in simulation models.

4.2 Queueing and Staffing

The major workforce management systems use the Erlang-C model for staffing, which is the basic M/M/c queueing model with the following assumptions:

- Poisson arrival process
- Exponential service times
- Multiple servers (c) in parallel with a single queue
- All servers are identical, i.e. same distribution (exponential) with the same mean
- Service is FCFS and consists of a single phase
- Unlimited queue length
- No balking or reneging (abandonments)

The M/M/c queueing model is reasonably robust so minor departures from these assumptions for the "one size fits all" call center are not all that significant, the key concern being the exponential assumption. However, in today's technology driven environment, many of these assumptions are invalid and continued use of the Erlang model results in varying degrees of over staffing. Additionally, this model cannot be used for all of the "what if" scenarios that address call handling strategies, call center designs, and call routing options. Simulation is a far superior modeling approach that overcomes many of the difficulties of analytical models and associated assumptions. Another key advantage of simulation for queueing and staffing is that it provides information on variability and extremes, unlike the queueing models that mostly provide averages for key performance measures (Klungle 1998).

4.3 Workforce Scheduling

This module is the main strength of workforce management systems. Given forecasts, desired service levels and associated staffing levels, and operating rules (shift schedules, breaks, lunches periods, meetings, vacations, productivity levels, etc.) staff schedules are developed based on various optimization criteria such as cost and/or revenue or a balance between over and under staffing.

4.4 Analytical Modeling

A workforce management system is like a large spreadsheet in that it can analyze scenarios by manipulating a series of inputs through a static model to produce a desired output. These systems along with queueing models can quickly provide reasonable approximations for mid to long term planning scenarios. Some examples would include staffing approximations for budgeting purposes, and "what if" analyses of changes to quality standards. Building simulation models for these purposes not only takes longer at times but also may not provide any additional significant or accurate information. On the other hand, many scenarios may have to be generated for a workforce management system whereas a single simulation model could answer all of the questions and issues under consideration.

5 SIMULATION

5.1 When to Simulate

There are a number of reasons for using simulation versus analytical models that deal with the deficiencies of the models themselves such as:

- Analytical models not available
- Existing analytical models are too complex
- Static results of analytical models are insufficient
- Analytical models only provide averages, not variability and extremes
- Analytical models cannot identify process bottlenecks or recommend design changes
- Analytical models often cannot provide sufficient detail nor identify interactions
- Animation is a better method of demonstrating results to management

5.2 Problems with Erlang-C

As previously noted, there are many deficiencies with Erlang-C and the classic queueing models. With the movement towards skill-based routing of calls due to advances in technology, Erlang-C is basically outdated for many applications since it assumes that agents have a single skill and there is no call priority (FCFS). Even with partial cross training, using Erlang-C can cause considerable overstaffing. The problem is compounded when it comes to scheduling the workforce. With skills-based scheduling, the call types an agent will handle are dependent upon both other schedules and upon call routing rules. With these complexities, simulation has become the best alternative for developing operational strategies for call centers.

6 AN INSURANCE CLAIMS APPLICATION

6.1 Background

This simulation project was the result of a re-engineering study of the entire insurance claim handling process, including the establishment and operation of additional call centers. Under the initial re-engineered environment, low service levels and high abandonment rates were the general rule leading to a significant decrease in member satisfaction levels.

6.2 Alternatives

The traditional approach to resolving this problem would simply be to add staff and this was the recommendation of call center management. However, a possible source of some of these service level problems is in the routing of incoming calls to the claim representatives. Thus, one key concept to be tested is that of organizing calls by length versus the more traditional approach where calls are grouped by functional type. In scheduling theory, a method of maximizing throughput is to use the SPT rule (Shortest Processing Time). This is basically the same concept used by retail establishments with express lines for "10 items or less, cash only". Additionally, as learned from queueing theory, high variability in call processing times causes longer queues and longer waits.

In call center management, this SPT concept can be achieved by the creation of a short-duration gate. However, a key issue with this approach is the higher skill level required of some claims representatives since a wider variety of call types must be handled. Thus, if the SPT concept were to be implemented, additional training (4-6 weeks in some cases) would be necessary.

6.3 Simulation

Since workforce management systems are not capable of modeling a wide range of scenarios, simulation becomes the tool of choice. Models were created for both the current and proposed operations. Call arrival patterns follow a Poisson process with the mean rate dependent on the time of day and day of week. Call lengths and after-call work time generally follow a gamma distribution with the two parameters being dependent on the type of call. Verification of the models included a comparison of:

- Simulated call volumes to Rockwell data (by time of day and day of week)
- Staffing levels to current call center staff (vacations, breaks/lunch, and training schedules)
- Simulated service levels to Rockwell reported service levels
- Abandon rates to Rockwell reported abandon rates (by time of day)
- Call handling times to Rockwell phone reports, prior pilot studies, and claim quality surveys
- Work time to Rockwell phone reports
- Call breakdowns to Claims Processing System reports (by type and time of day)

6.4 The Success

Results of the simulation models provided insights into two key areas, the first being the improvement in service without adding staff and the second being the identification of other process improvement opportunities. The short duration gate provided only slight improvements in the service level measurement criteria (percent of calls answered in a specified time period) but significant improvement in queue length, average waiting time, and abandonment rate as shown below in Table 1.

Table 1: Performance Measures for Current and Proposed Designs

	Model	
Performance Measure	Current	New Gate
Abandon Rate	20%	15%
Service Level (95/60)		
Gate A	43%	41%
Gate B	69%	70%
Gate C	67%	69%
New Gate		60%
No. in Queue		
Gate A	2.74	0.84
Gate B	0.27	0.23
Gate C	0.21	0.20
New Gate		1.29

Additionally, as new call centers open and more staff becomes available, the impact of this short duration gate will be somewhat greater. To meet the desired service levels in the call center, additional staff is still required, but the additional number required is about 20% less if the SPT rule is used.

6.5 The Failure

Despite the significant cost savings available from the SPT call-handling rule, management decided not to pursue this alternative. A key issue was the difficulty in separating the bimodal distribution of inquiry type calls into its two components, namely; a large mass of short duration calls with small variability, and a much smaller mass of long duration calls with high variability. It was felt that the additional training required along with the increased operational complexity basically offset any gains in efficiency. Thus the more traditional solution of staff additions was implemented. However, all was not lost. Two process improvement suggestions resulting from the sensitivity analysis were identified and investigated as follow-on projects. First, what approaches were available to reduce after-call work time for the "Generalist" category, and second, what would be the impact of decreasing the number of inquiry calls if certain actions were taken that could increase the average length of calls. Improvements were achieved in both areas.

7 SUMMARY AND CONCLUSIONS

This paper has reviewed the call center management process and key performance measures used to evaluate call center effectiveness. The use of workforce management systems was then addressed, highlighting some of the strengths and weaknesses of these systems and analytical models in general. The advantages of simulation as an analysis tool for call centers was demonstrated using an insurance claim application. The technical aspects of the project were highly successful while the implementation was not pursued. From the comparative analysis and example application some important conclusions can be drawn.

7.1 The Call Center Process

- Forecasting is the most critical step in the call center management process
- Analytical models and specifically Erlang-C
  - ⇒ are often sufficient for general planning purposes
  - ⇒ are generally sufficient for the traditional "one-size-fits-all" call center if the assumptions are reasonably met
  - ⇒ have a tendency to overstaff even the traditional call center as key assumptions are often violated
  - ⇒ are becoming an obsolete methodology as communications and call center technology advance

7.2 Simulation

- Simulation is a superior analysis tool when
  - ⇒ dealing with detailed operations
  - ⇒ accuracy is important
  - ⇒ demand variability is high
  - ⇒ animation is needed to sell concepts to management
  - ⇒ more than just averages are needed to understand the implications of process changes
  - ⇒ process bottleneck are to be identified
- Simulation can be used to experiment with new designs or policies without disruption to current operations, allowing call centers to prepare for
  - ⇒ changes in call volumes
  - ⇒ introduction of new products and services
  - ⇒ process improvements
  - ⇒ revisions to service level goals

7.3 Workforce Management Systems

- Workforce management systems are a mixed bag, working well for scheduling and as an assist to the entire call center management process, but limited in modeling and forecasting capability, and in the ability to deal accurately with advancing technologies such as skill-based routing.

7.4 Claims Application

- It requires more than good simulation results to sell a solution.
- Operational complexity is a key management concern.
- Secondary results of simulation projects can be quite beneficial.

8 FUTURE

When considering the future role of simulation in the call center management process, specialized call center simulation software becomes a consideration. Currently these are standalone products that greatly simplify the call center simulation process by significantly reducing the development time and by allowing for more flexibility in modeling. However, these products are an intermediate step in the entire process of making simulation a standard module in workforce management systems. With current and forthcoming technologies leading the change to more skill-based call routing, the entire call center management process is becoming much more complex. Current workforce management systems are not capable of dealing with this complexity. They will have to be redesigned to deal with these new realities. As part of this redesign process, simulation modules will have to be included as part of the overall solution.

REFERENCES

Klungle, R., and J. Maluchnik. 1997. The role of simulation in call center management. In *Proceedings of the 1997 MSUG Conference.*

Klungle R., and J. Maluchnik. 1998. Call center forecasting at AAA Michigan. *The Journal of Business Forecasting* 20 (4): 8-13.

Klungle, R., 1998. Forecasting and staffing in call centers: principles, procedures, perceptions, pitfalls, and practicalities. In *Conference on Workforce Management for Call Centers,* IQPC, Dallas, September, 1998.

AUTHOR BIOGRAPHY

ROGER G. KLUNGLE is manager of Business Operations Analysis at AAA Michigan and an adjunct professor of Industrial and Manufacturing Systems Engineering at the University of Michigan-Dearborn. He received a B.S. degree in Industrial Engineering from the University of Michigan, and a M.S. and Ph.D. in operations research from George Washington University. He is a member of IIE, ASQC, SCS, MSUG, and IBF. His primary interests are in business applications of data analysis and simulation. Prior to his current employment in the insurance industry, he worked in the energy and transportation industries, and for the Department of Defense.