

Designing a Telephone Call Center with Impatient Customers

Research Thesis
Submitted in Partial Fulfillment of the Requirements
for the degree of Master of Science
in Operations Research and Systems Analysis

OFER GARNETT

Submitted to the Senate of the Technion - Israel Institute of Technology

Adar 5758

Haifa

March 1998

The research was carried out under the supervision of Prof. Avishai Mandelbaum in the faculty of Industrial Engineering & Management.

I am grateful to Prof. Mandelbaum for his guidance during this research, his patience and assistance. I would also like to thank Dr. Marty Reiman of Bell Labs for his help.

The generous financial help of the Technion Graduate School is gratefully acknowledged.

Table of Contents

Introduction	4
1 The analytic model	9
1.1 Description as a Birth and Death process	11
1.2 Alternative description as a two-node network	12
1.3 Elementary performance measures - waiting and abandonment probabilities	13
1.4 Abandonment probability - usefull results	15
2 Literature review	19
3 Exact calculation of performance measures in an $M/M/N/B+M$ system (abandonment rate θ)	27
4 Approximations	32
5 Asymptotic behavior of the abandonment probability	35
6 Asymptotic behavior in Heavy Traffic	38
6.1 Abandonment and waiting probabilities ($\theta_N \propto N^\gamma$)	41
6.1.1 $\gamma = 0$	41
6.1.2 $\gamma > 0$	44
6.1.3 $\gamma < 0$	47
6.2 Weak convergence of the stationary distributions $\{q_N(\infty)\}$	50
6.3 Weak convergence of the processes $\{q_N\}$	56

Table of contents continued:

6.4	Approximating the stationary distribution - conclusion	70
7	Designing a service center - discussion	73
7.1	Examining quality of the approximations	75
7.1.1	Dependence on the number of servers	76
7.1.2	Application	79
7.1.3	"Extreme" abandonment rates	82
7.2	Rules of thumb	84
7.3	Service-level measures	87
	Conclusion / Future research	89
	Appendix: MATLAB code for exact calculations of performance measures	96

List of Figures

0.1	Schematic model of a telephone call center	5
1.1	$\{Q(t), t \geq 0\}$ - Transition diagram	11
4.1	Approximating the process $\{q_N(t), t \geq 0\}$	33
7.1	Dependence of $P\{Wait > 0\}$ on θ in an $M(95)/M(1)/100 + M$ system . .	81
7.2	$P\{Ab\}$ in an $M(95)/M(1)/100 + M$ system with "slow" abandonment rates	82
7.3	$P\{Ab Wait > 0\}$ in an $M(95)/M(1)/100 + M$ system with "slow" abandonment rates	83
7.4	$P\{Ab Wait > 0\}$ in an $M(95)/M(1)/100 + M$ system with "fast" abandonment rates	84

List of Tables

3.1	Performance measures of the form $E[f(V, R)]$	28
6.1	Results for asymptotic behavior in HW regime	42
7.1	Asymptotic behavior of the waiting probability in HW regime	76
7.2	Minimum number of servers to achieve 0.05 accuracy	77
7.3	$\gamma = 0$ - Minimum number of servers to achieve 0.01 accuracy	79

Abstract

During the recent decades there has been an explosive growth in the number of companies providing services via telephone, and in the variety of telephone services provided. In this work we focus on telephone call centers that provide information or business services, customer help desks, and telemarketing centers. Telephone call centers of this type are large (number-of-agents-wise), relative to other service centers.

The overall challenge in designing and managing a service center is to achieve a balance between efficiency and service level. Service via the telephone requires a relatively short response time (seconds), while the number of daily calls to a large telephone call center can reach tens of thousands. In these circumstances, one must rely on analytical models to achieve the sought-after balance. A common model for the analysis of telephone call centers with many servers is the $M/M/N/B$ model, but it lacks an important feature: in service centers, particularly service via telephone, customers tend to be impatient, therefore customers waiting in queue might decide to leave (abandon) before their service begins.

It turns out that the design of a call center using results from the analysis of an $M/M/N/B$ model (without abandonment) may cause a significant excess in server allocation. The abandonment of a client hurts the level of service provided, but by abandoning he is also relieving the workload in the system and improving the service level of the other customers. Therefore, it is important to equip the call center's designer/manager with tools which will enable the assesment of the impact of abandonment on service level and efficiency. Abandonment in the $M/M/N/B$ model also adds a "dimension" to performance measures, and the call center's designer/manager can account for "abandoning" customers in addition to "waiting" customers.

We have analyzed the simplest case, in which the customers' patience is modeled by an exponential random variable, independent of the arrival and service processes, and the state of the system. Assuming independence on the state of the system is appropriate for

service via the telephone, where the queues are "invisible" in the sense that customers typically have no information about the state of the system (in contrast, for example, to a bank, in which the customer sees the queue which typically effects behavior). For an $M/M/N/B$ model with "exponential" abandonment we use the notation $M/M/N/B+M$. We introduce two schemes for exact calculation of performance measures of the $M/M/N/B+M$ model in steady state. We continue by focusing on the asymptotic analysis of an $M/M/N+M$ system, in a regime appropriate for large service centers (many servers, high efficiency, high service level). Here we assume that the system's waiting capacity is unlimited. (This assumption coincides with the prevalent policy of avoiding busy-signals by allocating a large number of trunks for customers awaiting service). Our asymptotic analysis was performed at two levels:

1. For two elementary performance measures - fraction of customers required to wait in queue and fraction of customers who abandon before their service begins.
2. The stationary distribution of the stochastic process of the number of customers in the system. Here we used two different approaches - finding the weak limit of a series of stationary distributions, and finding the stationary distribution of the weak limit of a series of stochastic processes. The results in both cases coincide.

In both cases the asymptotic behavior depends on the values of the system's parameters, and their interrelations. Our primary goal is to use the asymptotic results as approximations. It is possible to determine which values of the parameters lead to the most "successful" approximations, and this defines a "sub-model" which is most suitable for the analysis of telephone call centers with impatient customers. Following the asymptotic behavior of this sub-model, we derive approximations for a variety of performance measures and propose "rules of thumb" for the design of a telephone call center.