A Framework for Balanced Service and Cross–Selling by Using Queuing Science

Ayhan Demiriz

Department of Industrial Engineering,
Sakarya University,
Sakarya, 54187 Turkey
ademiriz@sakarya.edu.tr
+90 264 295-5674

Ufuk KULA

Department of Industrial Engineering,
Sakarya University,
Sakarya, 54187 Turkey
ukula@sakarya.edu.tr
+90 264 295-5665

Nevra Akbilek

Department of Industrial Engineering,
Sakarya University,
Sakarya, 54187 Turkey
nakbilek@sakarya.edu.tr
+90 264 295-5686

Abstract:

Call centers are complex contact centers to handle large volume of inbound, outbound or both types of calls depending on the business purpose. Call centers assume the role of the primary contact medium for many companies from a wide range of industries with their customers or clients. Despite of being seen traditionally as adding cost to the companies' bottom lines, call centers are now viewed by many companies to turn a service request into an opportunity to sell additional products and services. This sales attempt is called cross-selling. The opportunity to generate profit from an existing customer-base is a key factor for a successful call center. This paper introduces a framework for balancing cross-selling and service activities in a call center setup from a queuing science point of view. The main goal of this study is to introduce a framework to maximize a call center's performance without degrading the service quality. Our framework is based on the usage of real-time queue characteristics, customer profile information and server-skill set information from a cross-sell point of view.

Keywords: call center; cross-selling; contact center; queuing theory; skill-based routing, call routing.

1. Introduction

Call centers also known as customer service centers are becoming more prevalent means to handle customer service request in a variety industries ranging from financial service to retail companies. Call center industry is expanding worldwide in rapid face in terms of workforce and economic. For example, it is estimated that 3% of workforce both in the UK and USA is employed by call center industry with 20% growth rate (Koole and Mandelbaum, 2001). In addition, outsourcing activities in developed countries create significant amount of jobs in developing countries like India.

Queuing theory concerns with the mathematical explanations and analysis of stochastic systems from congestion point of view. On the other hand, as mentioned in (Brown et al., 2005), queuing science deals with to validate and calibrate queuing-theoretic models by extensive statistical data analysis on real systems. Thus, queuing science can be considered as the empirical and practical answer to the operational and managerial problems in the call center industry. Nevertheless, many design decisions are made based on the theoretical results of the queuing theory.

Johnson and Seymour studied a retail bank before and after a cross-sell program initiated across its branches. Their study is the one of earliest analysis of cross-sell efforts in traditional business environment (Johnson and Seymour, 1985). Similar to their work, Akşin and Harker studied identifying some of the effects of increased sales activities on customer service in phone centers (Akşin and Harker, 1999). The primary goal in this particular study was to ascertain possible effects of cross-sell effort under different organizational scenarios. Akşin and Harker applied the framework proposed in their study to a retail bank call center. For a successful entering customer, the expected revenue is deterministic but dependent on the service type. They showed that appropriate managerial actions can ameliorate the congestive effects of cross-selling (Akşin and Harker, 1999).

(Gans et al., 2003) provides a tutorial to the operation and management issues in telephone centers. Koole and Mandelbaum prepared a survey as an introduction to the queuing models of call centers (Koole and Mandelbaum, 2001). Güneş and Akşin segmented customer base into high and low value customers and analyzed cross-selling attempts to one or both customer segments (Güneş and Akşin, 2004). Örmeci and Akşin modeled revenue generation and congestion in a call center as a dynamic framework (Örmeci and Akşin, 2005). We model our cross-selling problem as a queuing system for optimal control.

In this paper we introduce a framework for balancing cross-selling and service activities in a call center setup from a queuing science point of view. The main goal of this study is to introduce a framework to maximize a call center's performance without degrading the service quality. Our framework is based on the usage of real-time queue characteristics, customer profile information and server-skill set information from a cross-sell point of view. In addition, we propose a M/M/1 based solution for the problem of finding the optimum threshold (p*) for the cross-selling in the presence of regular service. For this purpose, we device a threshold function w.r.t. the number of customers in the call center system. We present results from this model under various scenarios such as traffic intensity, likelihood of purchasing, and initial cross-selling threshold. We compare the results of this model with the results of the full-policy model proposed in (Byers and So, 2007) under equivalent parameters experimental setup.

In the remaining of the paper, we give a short explanation of the call center environment and discuss its major components in Section 2. We then give a summary of common theoretic queuing models in Section 3. Section 4 introduces our framework proposed in this paper. We develop practical implementation of our framework as an M/M/1 model in Section 5. Section 6 reports our experimental results on M/M/1 model and its comparisons with baseline models. We also report some results from our implementation of the full-policy model introduced in

(Byers and So, 2007). We then conclude our paper with a conclusion and the future work section.

2. What is a Call Center?

Call centers are modern service networks in which customer service representatives (CSRs) provide services to customers via telephones and computers. Call centers can be categorized according to their functionalities (help desk, emergency, tele-marketing, information providers, etc.), agent characteristics (low-skilled, high-skilled etc.), their sizes (number of agent seats) and more (Koole and Mandelbaum, (2002)). A main characteristic of a call center is whether it handles inbound or outbound traffic. Our focus here is on inbound (incoming) call centers where the calls are initiated by the customers.

Call **(contact)** centers are technology-intensive operations. However, 70% or more of their operating costs are stem from human resources (Koole and Mandelbaum, (2002)). Well-run call centers maximize the expected operation profit without sacrificing the service quality. Successful call centers adhere to the agent efficiency through well balanced call routing based on queue characteristics (service quality) and customer profile information. Call centers can be viewed, naturally and usefully, as queuing systems see Figure 1.. In a queuing model of a call center, the customers are callers, servers (resources) are telephone agents (operators) or communication equipment, and tele-queues consist of callers that await service by a system resource. Our model is based on a single waiting line, multi-server and finite queuing model that works according to FCFS discipline.

Figure 1

Modern call center is often more complicated queuing networks. One of the early breakthroughs in call center technology is the PABX's (Private Automatic Branch Exchanges, or simply PBX), the telephone exchanges within companies. A PBX connects, the public telephone network to telephones within the call centers where are staffed by telephone agents, often called CSR (Customer Service Representatives). In between the PBX and the agents, there is the ACD (Automatic Call Distribution) switch, whose role is to distribute calls according to skill and idle among agents. ACD is also the archival collection of operational data, which is of first importance research for call center (Koole and Mandelbaum, (2002)).

Most call centers support Interactive Voice Response (IVR) units, or Voice Response Units (VRU's), which are the industrial versions of answering machines, including the possibilities of interactions. But more generally, a current trend is the extension of the call center into a contact center. Telephone service is enhanced by some additional multi-media customer-contact channels such as e-mail, fax, internet or chat (in that order of prevalence). The customer can be often automatically identified by the PBX, using Automatic Number identification (ANI). Computer Telephony Integration, (CTI) is then used to search for customer's history file. CTI and ANI are together used to identify the potential customer, for example, cross-selling opportunities and, thus routing of the call to an appropriately skilled agent. IVR and CTI can be used for reporting purposes.

3. Call Centers from a Queuing Theory Perspective

Queues (or waiting lines) are an unavoidable component of modern life. We are required to stand physically in queues in grocery stores, banks, department stores, amusement parks, movie theaters etc. Although we do not necessarily like standing in a queue, we appreciate the fairness that it imposes. Typically, a queuing system consist of a stream of customers (humans, finished goods, messages) that arrive at a service facility, get served according a given service discipline, and then depart. In practice we are interested in designing a queuing system, namely, its capacity, number of servers, service discipline, etc.

From a queuing theoretic point of view, call centers can simply be modeled as M/M/s, sometimes referred to as Erlang-C. In this s server model, arrivals follow a Poisson process and service times are exponentially distributed. This simple model is very restrictive besides the assumptions of the arrivals and service times are Markovian; it does not acknowledge customer impatience, abandonment behavior, customers' heterogeneity, or servers' skill levels (Brown et al., 2005). To provide a more realistic model, (Garnett et al., 2002) introduces M/M/s+M model which is also called Erlang-A to approximate the time to abandonment a call by an exponential distribution. '+M' denotes again Markovian inter-arrival times. Although customer patience should be properly included in a call center model, we leave the further discussions for the future work.

Another simple model is M/M/s/FCFS/K/∞ which considers limited capacity in the waiting line which is realistic. We briefly explain this model. Assume that $\{X_K(t), t \geq 0\}$ is a birth and death process on $\{0, 1, \ldots, K\}$ with birth parameters $\lambda_i = \lambda$ for $0 \leq i \leq K-1$ and death parameters $\mu_i = \min(i, s)\mu$ for $0 \leq i \leq K$. Characteristics of M/M/s/FCFS/K/∞ queuing model then can be defined as follows (Kulkarni, 1999).

$$\rho = \frac{\lambda}{s\mu}$$

$$\pi_i(K) = \pi_0(K)\rho_i \quad i = 0,1,2,\ldots,K,$$

$$\rho_i = \begin{cases} \dfrac{1}{i!}\left(\dfrac{\lambda}{\mu}\right)^i, & \text{if } 0 \leq i \leq s-1, \\[2ex] \dfrac{s^s}{s!}\rho^i, & \text{if } s \leq i \leq K, \end{cases}$$

Server idleness can be computed as

$$\pi_0(K) = \left[\sum_{i=0}^{s-1}\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i + \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \cdot \frac{1-\rho^{K-s+1}}{1-\rho}\right]^{-1}$$

Expected number of customers in the system is calculated by

$$L = 0\pi_0 + 1\pi_1 + 2\pi_2 + ... + n\pi_n$$

$$L = \sum_{n=0}^{K} n\pi_n = \sum_{n=0}^{K} i\pi_i(K)$$

Blocking probability for the finite

$$\pi_K(K) = \pi_0(K).\rho^K = \pi_0(K).\left(\frac{\lambda}{s\mu}\right)^K = \pi_0(K).\frac{s^s}{s!}\rho^K$$

For entering customers

$\pi_i(K) = 0$ for i $\geq$ K i.e. a customer cannot enter the system, if the system is full.

Expected waiting time for the entering customers is computed as

$W = \dfrac{L}{(\lambda(1-\pi_K(K)))}$, where $\lambda(1-\pi_K(K))$ is arrival rate of entering customers.

In the next section, we explain the fundamental components of our approach based on the work of Byers and So (Byers and So, 2004).

4. The Foundation of The Framework

We used birth and death processes to analyze the operations of a call center system with the assumptions that arrival and service times follow certain distributions. Our framework shares some similarities with (Byers and So, 2004), thus it is worthy to give summarize their full-policy model. Following the notation used in (Byers and So, 2004), we define that

$\lambda =$ Arrival rate

s= number of service agents(servers)

$\mu_X =$ Service rate for cross-selling on top of regular service

$\mu_S =$ Service rate for providing the regular service only

h=Holding cost for each customer per unit time

R= Revenue generated from each successful sale

C_{ss}=Cross-sell skill of servers

$\rho_1 = \lambda / s\mu_X$

$\rho_2 = \lambda / s\mu_S$

We assume in this study that the service time distributions for both providing cross-selling with the regular service and only regular service are both Markovian (exponential), with rates μ_X and μ_S respectively. We also assume that $\mu_X < \mu_S$ and $\rho_2 < 1$ throughout this study.

In (Byers and So, 2004), the probability distribution of a successful cross-sell for existing customer base is assumed to be uniformly distributed in $[q-\in, q+\in]$, where $0 \leq q-\in \leq q+\in \leq 1$. In other words, q is the long-run average probability that a random customer will make a purchase, and $\in$ is a measure of variability of successful cross-sell across the existing customer base. In our study, we employ the same assumptions regarding the customer purchase probability distribution, however different types of distributional assumptions are out of scope of this paper and are left for the future work.

The main goal of this study is to propose a framework to decide whether to cross-sell or not cross-sell by considering the system congestion information (queue length), customer profile information (likelihood to purchase) and cross-sell skill of server information. We consider cross-sell effort to an opportunity management for maximizing the expected operating profit in the system, which is equal to the expected revenue for successful cross-sell minus expected customer holding costs as used in general queuing theory.

We incorporate queue characteristics such as the length, customer profile and server cross-sell skill information to control cross–sell opportunity to an individual customer. All customers are separated into two segment based on the estimated probability of a successful sale to customers in Byers and So, (2004). The top $p\ (0 \leq p \leq 1)$ proportion of customers has the highest probability of successful cross-sell as high-value customers. The remaining customers belong to low value group. A threshold value is used for cross-sell to an individual customer if and only if the total number of customers in the system is less than or equal to some pre-determined threshold when that individual customer is being served. They assume $n_h \geq 0$ and $n_l \geq 0$ to denote the respective thresholds used for high-value and low-value customers Byers and So, (2004). In our work, we propose a dynamic threshold which is only dependent on the number of customer in the system. Thus, our model is free of the parameters *p*, n_l, and n_h.

Under the exponential service times assumption, we use birth-death process to determine expected operating profit. All customers are assumed to be first-come first-served. Let S= $\{ (i,n): 0 \leq i; 0 \leq n \leq \max(i,s) \}$ denote the state space of underlying birth-death process, where *i* denotes the total number of customers in the system and *n* denotes the number of high-value customers currently by the servers. Let $\pi_{i,n}$ denote the steady-state probability at state *(i, n)*. We define $\mu_h(i)$ and $\mu_l(i)$ as the service rates for the high-value and low-value customers when there are *i* customers in the system, respectively.

$$\mu_h(i) = \mu_X, 0 \leq i \leq n_h$$
$$\mu_l(i) = \mu_{S,} i > n_h$$

$$\mu_l(i) = \mu_X, 0 \leq i \leq n_l$$
$$\mu_l(i) = \mu_{S,} i > n_l$$

$$\pi_{i,j} = 0, \qquad \text{for i < 0 or j < 0}$$

For $0 \leq i < s$ and $0 \leq n \leq i$

$$\{\lambda + n_h \mu_h(i) + (i-n)\mu_l(i) \} = p\lambda\pi_{i-1,n-1} + (1-p)\lambda\pi_{i-1,n} + (i+1-n)\mu_l(i)\pi_{i+1,n} + (n+1)\mu_h(i)\pi_{i+1,n+1}$$

For $i \geq s$ and $0 \leq n \leq s$,

$$\{\lambda + n\mu_h(i) + (s-n)\mu_l(i) \}\pi_{i,n} = p\lambda\pi_{i-1,n-1} + (1-p)\lambda\pi_{i-1,n} + (s+1-n)\mu_l(i)\pi_{i+1,n} + (n+1)\mu_h(i)\pi_{i+1,n+1}$$

The probability distribution of a successful cross-sell of all customers is assumed to be uniform $[q-\in, q+\in]$ where $0 \leq q \leq 1$. Hence, the probability distribution of purchase for high value customers is uniform $[q+(1-2p)\in, q+\in]$, which implies that the average probability of a purchase by any random high-value customer is equal to $[q+(1-p)\in]$. The arrival rate for high-value customers is equal to λp, the mean revenue per unit time because of cross-selling to high-value customers is equal to $R\lambda p[q+(1-p)\in]$. Similarly arrival rate of law value customers is equal to $\lambda(1-p)$and the expected reward per unit time because of successful cross-sell to low–value customers is equal to $R\lambda(1-p)[q-p\in]$. Thus, the expected operating profit is equal to (Byers and So, 2004)

$$F(p,n_l,n_h)=\lambda Rp[q+(1-p)\in]\sum_{i=0}^{n_h-1}\sum_{n=0}^{\min(i,s)}\pi_{i,n}+\lambda R(1-p)[q-p\in]\sum_{i=0}^{n_l-1}\sum_{n=0}^{\min(i,s)}\pi_{i,n}-h\sum_{i=1}^{\infty}\sum_{n=0}^{\min(i,s)}i,\pi_{i,n}.$$

We also assume for this study that the cross-sell skill of the server (C_{ss}) is multi-leveled e.g. Faster and Slower customer agents. This introduces a new index for the state space namely the level of cross-sell skill. We can devise a policy to route incoming calls to better skilled agents first if available otherwise route the calls to the lower level skilled agents. It should be noted that the skill levels are related to the individual agent's conversion rate i.e. successful cross-sell ability. In well-run call center, it is known that those more skilled agents have better conversion rates and faster cross-selling and service times. Then let S= $\{(i,n,j): 0\le i; 0\le n\le \max(i,s), 0\le j\}$ denote the state space of underlying birth-death process, where *i* denotes the total number of customers in the system, *n* denotes the number of high-value customers currently by the servers and *j* denotes the cross-selling skill level (in other words agent type). Then, let $\pi_{i,n,j}$ denote the steady-state probability at state *(i,n,j)*. Similar to above derivation, we can define different service, cross-selling rates for different type of customers and agents at the same time. This will obviously enlarge the state space, however it will result in a more realistic setup. By finding new steady state probabilities we can generate a new profit function in which we can differentiate different types of service rates and cross-selling revenues.

Our framework (see Figure 2) differs from prior work in many ways. First of all, we consider adding information about agent cross-sell skills (i.e. agent type). Moreover, the centerpiece of our framework is a new product recommender system (Demiriz, 2004) that works concurrently with an ACD. This new product recommender system is devised to optimize the cross-selling activities by considering the real-time queue characteristics, customer profile, contact history, and the cross-sell ability of the agents. In a call center environment, we might also consider adding individual sales targets of each the agent into our framework. But this needs a careful consideration and it is out of scope of our work.

Figure 2

In the next section we introduce M/M/1 based formulation of our framework. The back-bone in our approach is to utilize a dynamic *p* value depending on queue length. In other words, at each state we use a different threshold to specify the high level customers.

5. Dynamic Threshold for Cross-selling Under M/M/1 Assumptions

The arrivals are again assumed to be Poisson with parameter λ and service times are exponentially distributed. Depending on the decision about the customer, the service rate will be either with cross-selling (μ_X) or just the regular service (μ_S) rate. If the current customer falls into the top *p fraction* group then cross-selling attempt will be made by the customer service representative (agent). Thus the top *p fraction* of the customers can be considered as the high value customers and the rest of the customers can be considered as the low level customers.

The *p* value can be dynamically determined depending on the number of customers in the system. For example we can utilize the family of $p(n) = p^{an}$ functions where *a* could be 1, ½, 1/3, ¼ etc. Figure 3 depicts the different types of functions with the initial *p* value of 0.9. In this figure, *p* is a function of *n* (the number of customers in the system). We can call *p(n)* as the dynamic cross-selling threshold.

Figure 3.

For example if there are 5 customers in the system then according to $p(n) = p^n$ function approximately the top 60% of the customers will be treated as high value customers. When there are 10 customers in the system then approximately the top 31% of the customers will be treated as the high value customers. This can be done by simply ranking the customers according to some segmentation scheme. So we can actually utilize a segmentation scheme in our approach.

For the case of one customer in the system, since the arrival rate is λ and the dynamic cross-selling threshold is p_1, the expected service rate will be $p_1\mu_x + (1-p_1)\mu_S$ where μ_X and μ_S are cross-selling and regular service rates respectively. Similarly when there are n customers in the system, the service rate will be $p_n\mu_X + (1-p_n)\mu_S$.

Since the arrivals are Poisson process and the service times are exponentially distributed, this system can be considered as a birth and death process. We can find the steady state probabilities for this birth and death process as follows.

$$\pi_0 = \frac{1}{1+\sum_{n=1}^{\infty} c_n} \qquad \text{where} \quad c_n = \frac{\lambda_0\lambda_1.....\lambda_n}{\mu_1\mu_2.....\mu_n} \quad \text{and} \quad \mu_n = p_n\mu_X + (1-p_n)\mu_S$$

Figure 4.

The probability of being n customers in the system can be calculated as $\pi_n = c_n\pi_0$. For the birth and death process depicted in Figure 4 to be in steady state, the sufficient condition is that λ is less than the slowest service rate i.e. μ_X. Thus for the steady state distribution the sufficient condition is $\rho = \frac{\lambda}{\mu_X} < 1$.

When the cross-selling attempt is made to any customer, considering the successful cross-selling is uniformly distributed in $\left[q-\varepsilon, q+\varepsilon\right]$ then the average successful cross-selling rate will be q for any give customer. Thus the probability for the successful cross-selling for the top 50% of the customers will be greater than q, and for the bottom 50% of the customers will be less than q . In the case of n customers in the system, the success rate for the top *p(n)*% will be uniformly distributed in $\left[q+\varepsilon-2\varepsilon p(n), q+\varepsilon\right]$. Thus the expected (average) success rate will be $q+\left(1-p(n)\right)\varepsilon$.

5.1. The Objective Function

Assume that revenue generated from a successful cross-selling attempt is R. Since the arrivals follow Poisson process, the rate of n customers will be seen by the customers who arrive the system is the long-run probabilities (steady state) of being n customers in the system (PASTA property). Thus when call center is empty, the rate of incoming customers will be equal to long-run probability that the system is empty i.e. π_0. Therefore, the expected revenue for the arriving customer when the system is empty will be $R\pi_0 E[P(\text{Successful cross-selling, High value customer})]$. Since the events of the being a high value customer and having a successful cross-selling are independent, we simplify this expected value as follows.

$$R\pi_0 E[P(\text{Successful cross-selling})\text{P(High value customer)}]$$

$$= R\pi_0 p_1 E[P(\text{Successful cross-selling})]$$

$$= R\pi_0 p_1 (q+(1-p_1)\varepsilon)$$

Therefore, the expected revenue per unit time for the arriving customer when the system is empty is $\lambda R\pi_0 p_1 (q+(1-p_1)\varepsilon)$.

Similarly, the expected revenue generated from the arriving customers when there are i customers in the system is $\lambda R\pi_i p_{i+1} (q+(1-p_{i+1})\varepsilon)$.

If we assume the cost of waiting time per customer per unit time is *h*, the expected profit of the p* policy will be $\lambda R\sum_{i=0}^{\infty}\pi_i\left(q+\left(1-p_{i+1}\right)\varepsilon\right)p_{i+1}-hL$ where L is the expected number of customers in the system i.e. $L=\sum_{n=0}^{\infty}n\pi_n$.

6. Experimental Results

In this section, we conduct some experiments on our proposed M/M/1 dynamic *p* value (p* policy) model under various scenarios such as traffic intensity ($\rho=\frac{\lambda}{\mu_X}$), initial *p* value and customer variability (ε). We observe the change in the profit under these scenarios.

Figure 5

Figure 5 depicts that for all initial *p* values ($p(n)=p^n$) when the traffic intensity *ρ* increases, the expected profit first increases and then after a peak point it starts decreasing. There are two reasons for that. The first one is, when the intensity increases it will be less likely to have few customers in the system, thus agents will be more selective to cross-sell which will reduce the expected revenue. The second one is, once the system gets crowded with the increasing intensity, the expected costs will increase naturally by increasing waiting times.

It is evident from Figure 5 that the choice of initial *p* value (cross-selling threshold) have less impact on profit margin at lower traffic intensities than at the medium and high traffic intensities. This is due to the fact that regardless of the choice of *p(n)* values at lower traffic intensities, fewer number of customers receive cross-sell attempt. This is an important insight for the practical reasons. Because there is no need to calibrate the initial *p* values at lower traffic intensities. For the medium and high level traffic intensities, it is recommended to calibrate the initial p value (cross-selling threshold).

As seen in Figure 6, the effect of the variability of the cross-selling success (i.e. purchasing probability of the customer) is not noteworthy (insignificant). Therefore we can neglect the effect of the variability in the customer profiles in call centers working at lower traffic intensities. However, it is also evident that managing such variability is crucial for the call centers run at higher traffic intensities to increase the profit margin.

Figure 6

Figure 7 shows that call centers with medium level traffic intensities always operate at higher profit levels for all initial p values. This result supports the fact that there are fewer cross-selling attempts at lower traffic intensities due to the fewer customers and fewer cross-selling attempts at higher traffic intensities due to the increasingly selective attempts. Therefore call centers working at medium level traffic intensities are the best ones to offer cross-selling as reported in (Byers and So, 2007).

Figure 7

We can compare our results with baseline naïve models: a) Attempt each customer to cross-sell b) Never attempt to cross-sell and just offer the regular service. Second naïve model will never generate any revenue since the regular service does not generate any revenue. The total cost will be $-hL$. Obviously the expected L will be less than the systems with the cross-selling attempts. There maybe some extreme cases where this loss model would be comparable to the alternatives.

In the case of offering cross-sell to each customer, the expected profit will be $F = \lambda Rq - hL$. Certainly, the expected L would be larger than the selective cross-selling case. Therefore it may not be desirable to offer cross-selling to all customers.

We also run some experiments by implementing the full-policy model introduced in (Byers and So, 2007) to compare with our model and discuss similarities and differences. We report some results in Figure 8 and Figure 9. In these experiments, we change the traffic intensity ($\rho = \frac{\lambda}{\mu_X}$) from 0.2 to 0.95 by using the following parameter set (μ_S=2, R=10, h=1, p=0.5, q=0.5, ε=0.2, and n_h=10). We use n_l=1 in Figure 8 and n_l =5 in Figure 9.

Figure 8

Figure 9

As Figures 8 and 9 show, increasing n_l in (Byers and So, 2007) model has an increasing effect in the number of customers in the call center and a decreasing effect on the long-term call center profit. Since an high n_l means that the call center is less selective about the cross-sell attempted customers, from Figures 8 and 9 we see how important to be selective on cross-sell attempts could be. Our dynamic p value model takes this fact into the account and allows the call center to become more selective in its cross-sell efforts as the number of customers in the call center increases.

(Byers and So, 2007) classifies the call center customers into low and high value customers by assuming that top p fraction of the customer portfolio is high class customers. However, they further assume that both class of customers have identical purchasing likelihood and a customer in either class brings in the same cross sale revenue. This is not very realistic since customers in different classes usually will have quite different purchasing probability and/or will generate different cross-sale revenue. Our model allows us to incorporate different customer purchasing behavior into call center modeling since we may choose different purchasing likelihood for each *p(n)* fraction of customer portfolio defined by p^* policy.

7. Conclusion and The Future Work

We proposed a new framework for cross-sell efforts in a call center from a queuing science point of view by considering agent skill set as well compared to previous work in the literature. We also propose a new recommender system that works with automatic call distribution systems to route the call to appropriate agent at real-time. A recommender system fully integrated with ACDs will certainly need constant monitoring of the call center environment.

We believe that our M/M/1 p* policy is a modest starting point to eventually a generalized version where multi-server and limited space queues can be modeled more realistically. Our experimental results show similar results with an earlier model proposed in (Byers and So, 2007). However, we showed that our assumptions are more realistic and robust even for an M/M/1 call center environment.

Our very next work will be to establish theoretical foundations of our framework for a multi-server and limited space queue. One important aspect, we need to address is to model customer patience and abandonment behavior in real-time queues. As mentioned briefly, our starting point will be the assumption that the distribution of the time to abandonment is exponential.

On the other hand, the distributional assumptions on customer purchasing likelihood need to be generalized to further cases. In addition, we need to study the effects of different types of revenue schemes such as higher value customer may generate more revenue compared to lower value customers in a cross-selling transactions. Moreover, we need to study the cases where the regular services have a bottom-line revenue effects. That will be the case for a more interesting trade-off between cross-selling and regular service.

References:

Akşin, O.Z., and P.T. Harker, 1999, "To Sell or Not to Sell: Determining the Trade-offs Between Service and Sales in Retail Banking Phone Centers," Journal of Service Research, Vol. 2, pp. 19-33.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. 2005, "Statistical analysis of a telephone call center: a queuing science perspective". Journal of American Statistical Association. Vol. 100, No. 469, pp. 36-50.

Byers, R. E. and So K.C.,2004 "The value of Information-based Cross-sales Policies in Telephone Service Centers". Working Paper, University of California, Irvine, CA.

Byers, R. E., So, K.C. "A Mathematical Model for Evaluating Cross-Sales Policies in Telephone Service Center". Journal of Manufacturing and Service Operations Management Vol. 9, No. 1, Winter 2007, pp. 1-8.

Demiriz, A., 2004. Enhancing Product Recommender Systems on Sparse Binary Data. Journal of Data Mining and Knowledge Discovery, Volume 9, Issue 2, September 2004, Pages 147 – 170

Duxbury, D., Backhouse, R., Head, M., Lloyd, G., and Pilkington, J. 1999. Call centers in BT UK customer service. British Telecommunication Engineering, Vol. 18, pp.165-173

Johnson, Eugene M. and Daniel T. Seymour, 1985. “The impact of cross-selling on the Service Encounter in Retail Banking” in The Service Encounter.

Gans, N., G. Koole, G. And A. Mandelbaum, 2003, “Telephone Call Centers: Tutorial, Review and Research Prospects,” Journal of Manufacturing and Service Operations Management, Vol. 5, No.2, pp.79-141.

Garnett, O., Mandelbaum, A., and Reiman, M. (2002), “Designing a call-center with impatient customers”. Journal of Manufacturing and Service Operations Management. Vol. 4, No.3, Summer 2002, pp. 208-227.

Güneş, E. D. and R. Akşin, 2004,” Value Creation in Service Delivery: Relating Market Segmentation, Incentives and Operational Performance,” Journal of Manufacturing and Service Operations Management. Vol. 6, No. 4, Fall 2004, pp. 338-357

Koole, G. M., A. Mandelbaum. 2002. “Queueing models ofcall centers:An introduction”. *Annals of Operational Research.* 113 41–59.

Kulkarni, V. G., 1999. Modeling, analysis, design, and control of stochastic systems. Springer, NY, USA.

Örmeci, E.L., and Akşin, O.Z., 2004, “Revenue management through Dynamic Cross-Selling in Call Centers,”. Working Paper, Koc University, Turkey.

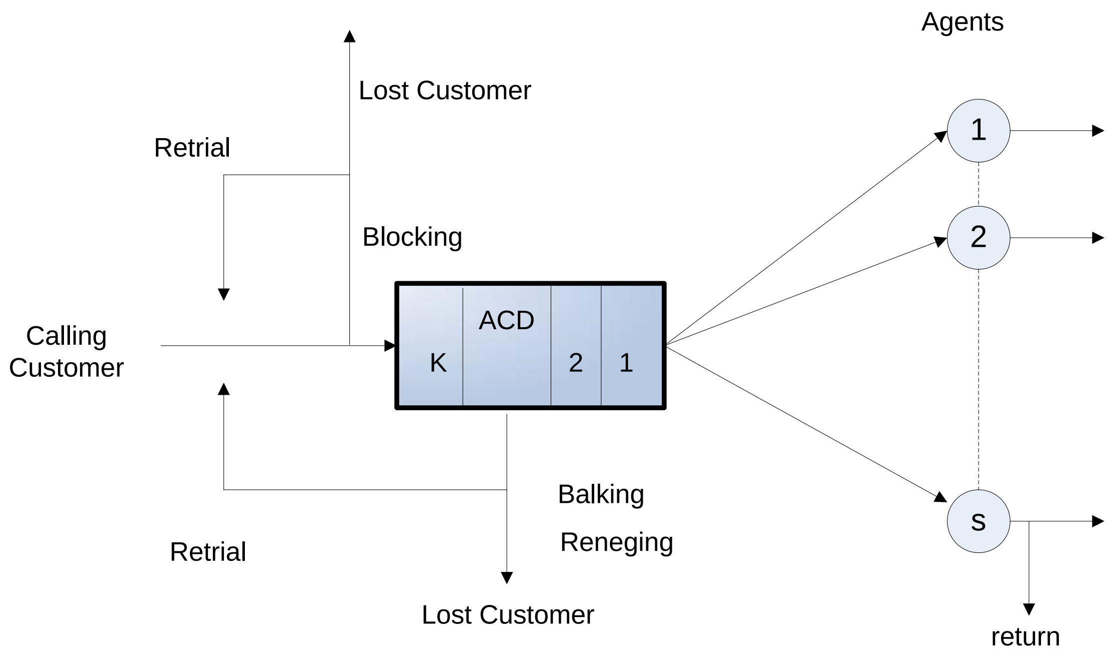

Figure 1. Operational scheme of a simple call

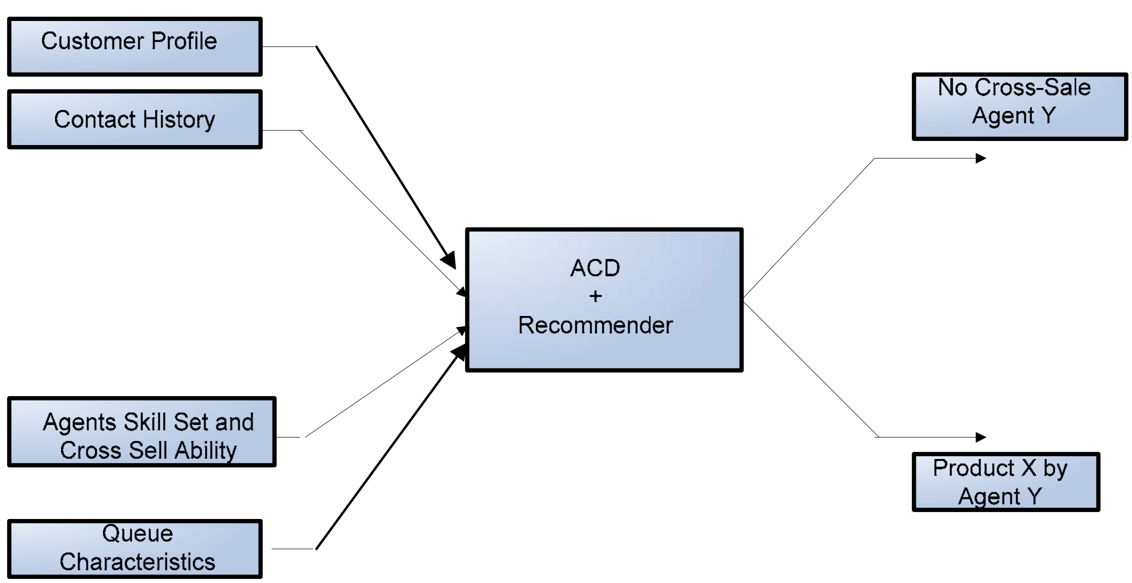

Figure 2. Proposed Framework

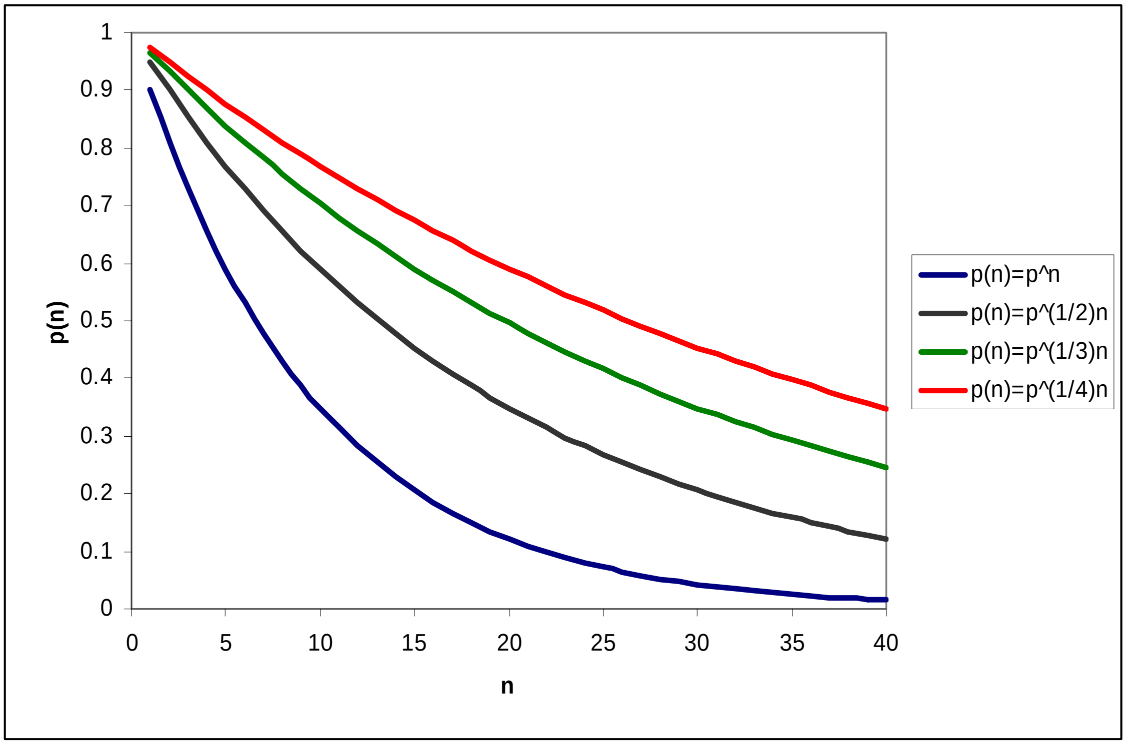

Figure 3. Dynamic *p* values vs. the number of customers in the system for different *p(n)* functions.

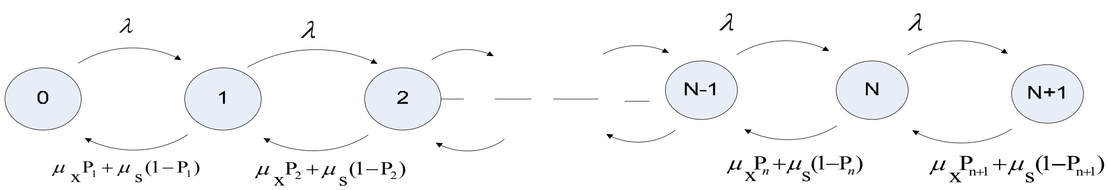

Figure 4. M/M/1 birth and death process under dynamic *p* values

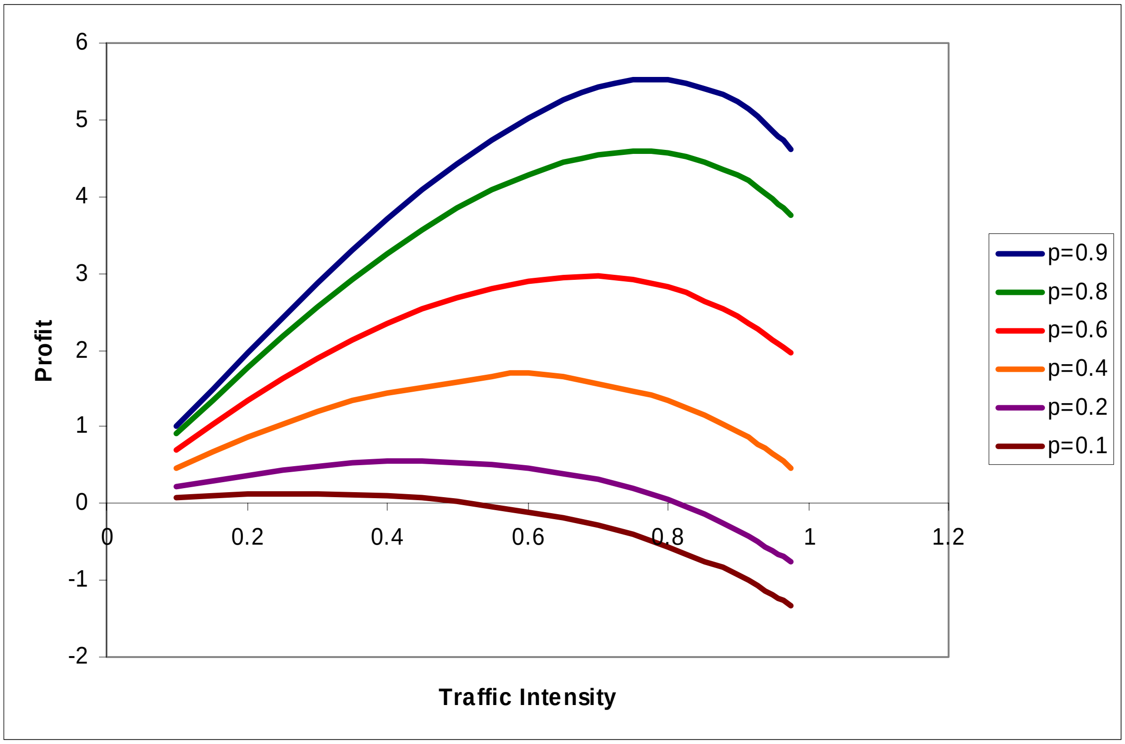

Figure 5. For the different initial p values, Profit vs. Traffic Intensity

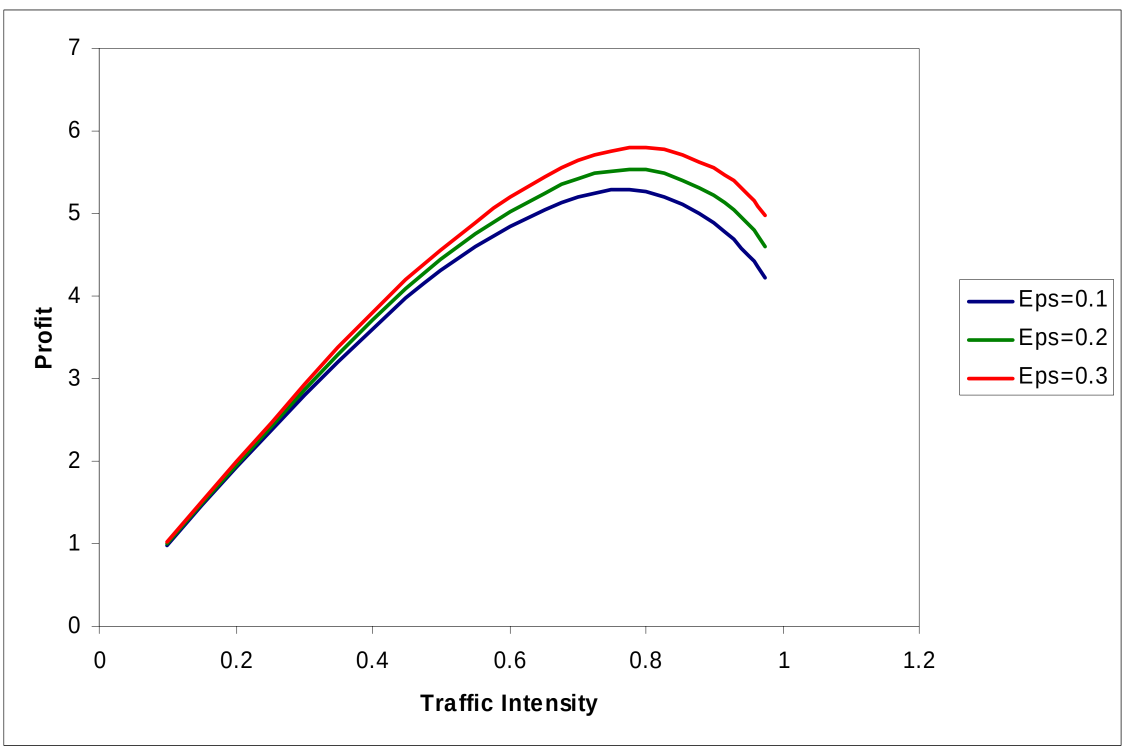

Figure 6. For the different ε values, Profit vs. Traffic Intensity

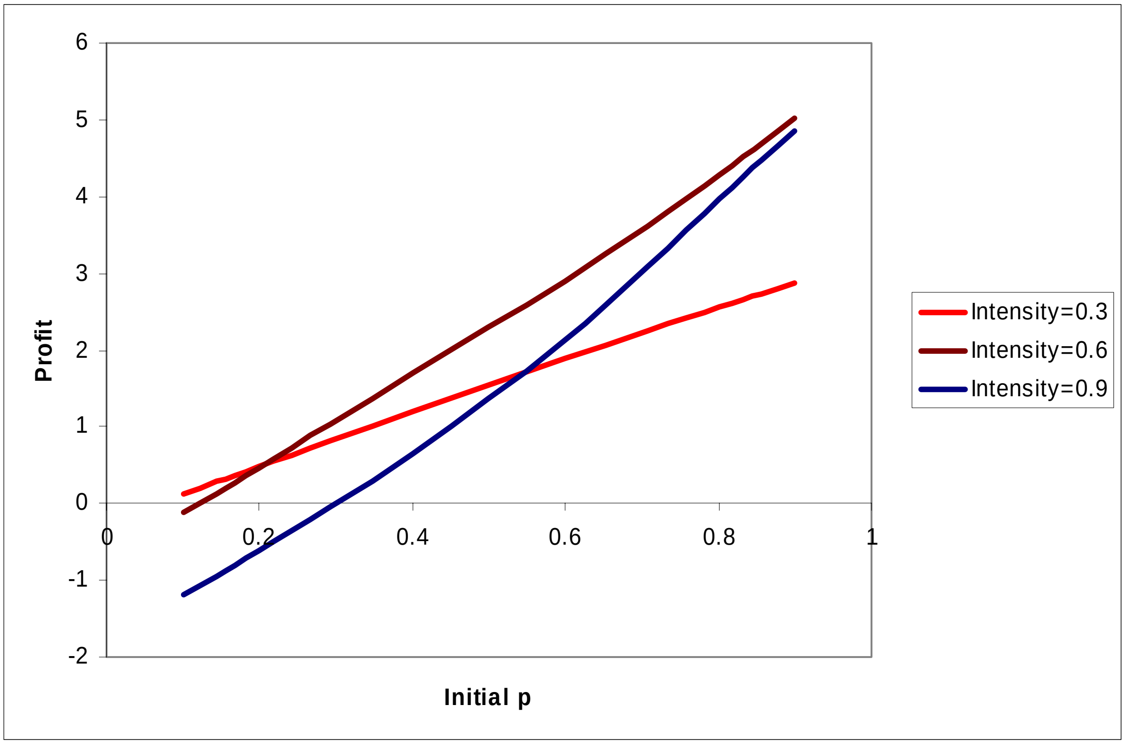

Figure 7. Under different traffic intensity, the affect of picking initial *p* values on profit

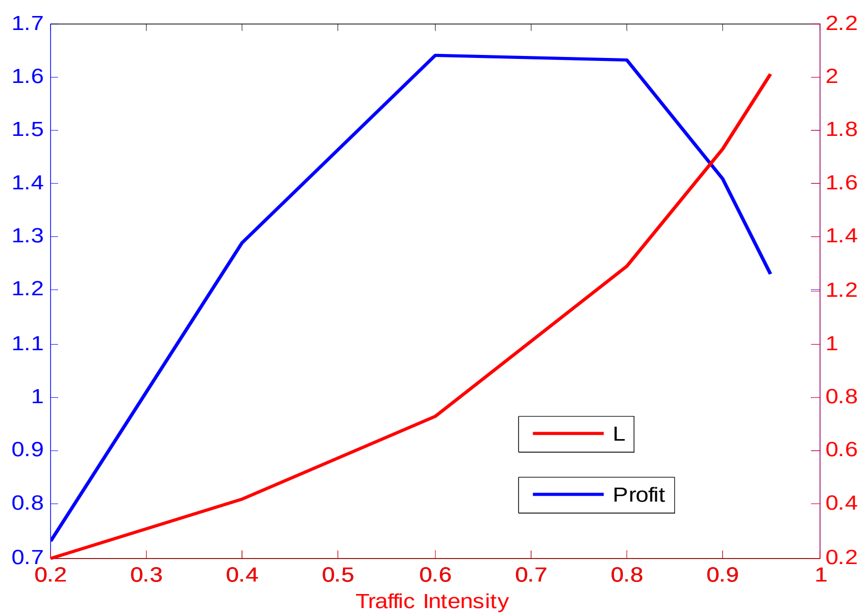

Figure 8. (Byers and So, 2007) model with n_l=1, n_h=10

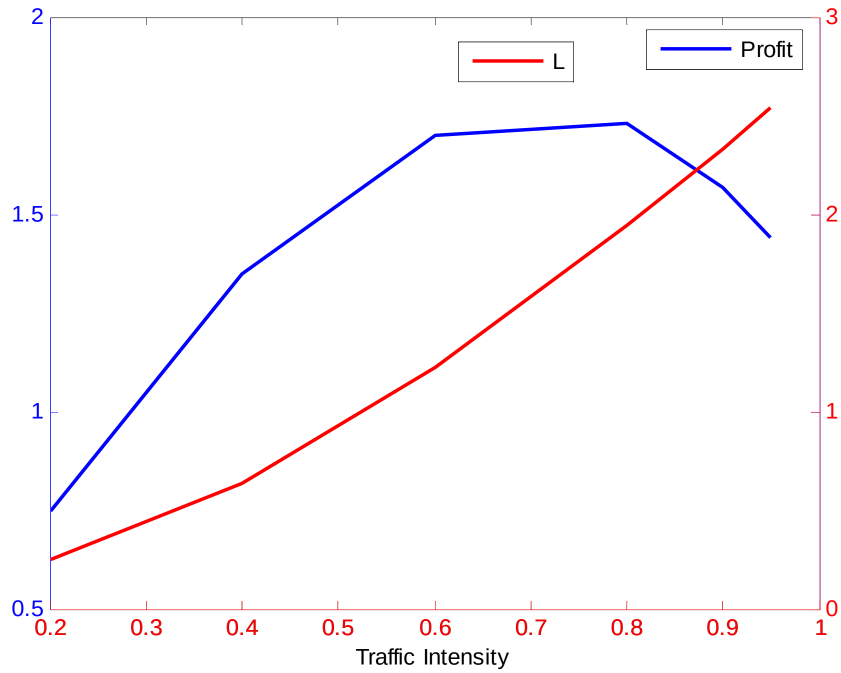

Figure 9. (Byers and So, 2007) model with n_l=5, n_h=10