Dynamic admission control in a call center with one shared and two dedicated service facilities

E. Lerzan Örmeci

Department of Industrial Engineering
Koç University
Rumeli Feneri Yolu
34450 Sarıyer, İstanbul, Turkey

email address: lormeci@ku.edu.tr
Tel: 90(212)338-1534
Fax: 90(212)338-1548

Abstract

Admission control of a call center is addressed through a Markovian loss system with two classes of calls and three stations, one dedicated for each class, and one shared station. Each class is identified by different revenues, service and arrival rates. We show that serving a call in its dedicated station, whenever possible, is optimal. For the shared station, we establish the existence of optimal monotone thresholds, and of preferred class(es) under certain conditions.

Acknowledgement: The author likes to thank Apostolos Burnetas (Case Western Reserve University, Cleveland) and Stephen Gilbert (University of Texas, Austin) for motivating this research.

The research was done while the author was at EURANDOM, Eindhoven, The Netherlands.
I. INTRODUCTION

In the last decades, call centers have been an effective and low cost customer service in a variety of industries such as financial services, airlines, hotels, and retail companies. Typically, call centers serve different classes of customers, each of which can be distinguished by a different profitability, volume of calls, and expectation for service. Call center operators often require different types of training to handle different classes of calls. However, because cross-training is expensive, it is necessary to manage the workforce training and call allocation carefully. In order to develop insight regarding the extent to which operators would be cross-trained, and how calls should be dynamically assigned to operators, we consider the problem of call admission in a system that serves two classes of calls using dedicated and shared facilities.

Consider a call center which serves two classes of calls, each of which requires different sets of skills from the operators of the center. At one extreme, the center can have two completely independent call centers, each dedicated to one of the classes, and at the other one it may consist of only one type of operators who are able to serve both classes. The former system suffers from either low utilization or low quality of service due to a loss in economies of scale, whereas the latter one has to face the high cost of cross-training many operators. Hence, we consider an intermediate model in which the call center consists of two dedicated stations, one for each class, and one shared station that can serve both classes.

The fraction of calls that is blocked due to busy signals or abandons the queue can have a drastic effect on system performance, (see [5]), which significantly influences the process of determining the required number of operators. To account for the effect of busy signals, we assume that the system has no waiting room so that calls are assumed to be lost if all servers are busy, or if the system chooses not to serve them.

Our model assumes that calls of each class arrive at the system according to a Poisson process, and demand exponential service times. Each class has different arrival and service rates, and generates different rewards. Moreover, the service rate for each class is different in the shared and dedicated stations. We do not allow preemption. Our objective is to derive the structure of dynamic admission policies that maximize the total expected discounted revenue over an infinite horizon as well as the long-run average revenue.

We show that the optimal policy accepts calls of each class to their dedicated stations whenever
that station has available servers. The optimal admission policy in the shared station, on the other hand, can be characterized in two ways: We prove the existence of a monotonic threshold policy. Moreover, we derive a sufficient condition for each class to guarantee that calls of that class are always admitted to the shared station whenever their dedicated station is full, and the shared station has at least one idle server. We call such a class a *preferred class*.

This paper is organized as follows: The next section presents a brief literature review on the related studies. Section III develops a Markov decision process (MDP) model for the system described above. Section IV characterizes the structure of optimal admission rules for the dedicated stations. In sections V and VI, we consider the admission control in the shared station: Section V presents sets of sufficient conditions for each class to be preferred, while section VI establishes the existence of an optimal monotonic threshold policy. In section VII, we illustrate and discuss our results via numerical examples. Finally, we conclude and point out possible future research in section VIII.

II. RELATED RESEARCH

Our work is related with the literature in two different areas: call centers due to our motivation, and admission control of loss systems due to our model.

The increasing use of call centers in different industries has generated recent research on these systems. Gans, Koole and Mandelbaum [5] give a comprehensive review for analytical models that support capacity management issues of call centers. Complexity of call centers provides a wide variety of different problems: Workforce management has become one of the concerns, see e.g., [2]. Computing performance measures of call centers is another important issue, addressed in [1] among others. Our work, on the other hand, is most closely related to the problem of routing different kinds of calls to servers with different skills. There are a number of canonical designs for skill-based routing (some of which are depicted in Figure 1) introduced in [7]. In V-design, two classes of calls are served by a single pool of cross-trained agents, variants of which are analyzed in a number of studies such as [4] and [6]. The N-design assuming fixed static priority policies is examined in [16] and [15], whereas Xu, Righter, and Shanthikumar [18] consider the optimal dynamic routing of such systems in a different context. In this paper, we address the dynamic control in an M-design with no waiting room.

Admission control is a main focus of research on loss networks, see Chapter 4 of [13] for a
A comprehensive review on the admission control problems of generalized stochastic knapsacks. More recently, Altman, Jiménez and Koole [3] and Koole [8] prove the existence of an optimal admission policy in a stochastic knapsack with Poisson arrivals and two classes of customers, that is characterized by acceptance thresholds, while Örmeci, Burnetas, and van der Wal [12] establish the monotonicity of these thresholds under some restrictive conditions. Örmeci et al. [12] and Savin, Cohen, Gans and Katalan [14] also introduce the notion of “preferred class” in stochastic knapsacks. In our formulation, the shared station can be viewed as a stochastic knapsack with two classes of customers: Our main contribution is to investigate the dependence of admission control in the shared station on the state of the dedicated stations. We show that existence of an optimal threshold policy with preferred class(es) in the shared station is not altered by the presence of the dedicated stations. Moreover, the thresholds are monotone in the number of calls in the dedicated stations. To the best of our knowledge, [9] is the only other study addressing admission control in a loss system with multiple stations.

III. MARKOV DECISION MODEL

We assume that class-$i$ calls arrive at the system according to a Poisson process with rate $\lambda_i$, $i = 1, 2$. Station $i$, having $c_i$ servers, is dedicated to calls of class $i$, $i = 1, 2$, whereas station 0 with
servers can be used by either of the classes. If a class-\(i\) call is served at station \(i\), it demands an exponential service with rate \(\mu_i\), and if it is served in station 0, then it takes an exponential amount of time with rate \(\mu_0\). We assume, without loss of generality, that class-1 calls require longer service times so that \(\mu_1 \leq \mu_2\) and \(\mu_1' \leq \mu_2'\). The operators in station 0 will be trained to be able to serve both classes, so we assume that these servers are slower than the “specialized” servers, i.e., \(\mu_j \leq \mu_j'\). Each finished class-\(i\) call brings a reward of \(r_i\) upon completion, regardless of the station it is served.

The original process of the system evolves in continuous time. All the interarrival times as well as the service times are exponential. Furthermore, we interpret discounting as exponential failures, i.e., the system closes down in an exponentially distributed time with rate \(\beta\) (for the equivalence of the process with discounting and the process without discounting but with an exponential deadline, see e.g., Walrand [17]). Hence, we can build a discrete time equivalent of this system by using uniformization (introduced by Lippman [10]): The maximum possible rate out of any state is finite, so we assume, using the appropriate time scale, \(\sum_{i=1,2}(\lambda_i + c_i \mu_i') + c \mu_2 + \beta = 1\). As a result, we observe the state of the system at each instant of a potential transition, so in every exponentially distributed time with rate 1, and the system changes its state with certain probabilities to be specified below.

Now we can define the state of the system as \((y; x) = (y_1, y_2; x_1, x_2)\), where \(y_i\) is the number of class-\(i\) calls in station \(i\), and \(x_i\) is the number of class-\(i\) calls in station 0. Let \(S\) be the set of all feasible states, i.e., \(S = \{(y; x): y_i \leq c_i, i = 1, 2; x_1 + x_2 \leq c\}\), so that if \(s = (s_1, s_2; s_3, s_4) \in S\), \(s_1\) and \(s_2\) are the number of calls in stations 1 and 2, respectively, whereas \(s_3\) and \(s_4\) are the number of class-1 and class-2 calls, respectively, in station 0.

Let \(u^n(y; x)\) be the maximal expected \(\beta\)-discounted reward for the system starting in state \((y; x)\) when \(n\) observation points remain in the horizon. Now we present the optimality equations, where \(e_j\) is the unit vector which has a 1 at the \(j\)th coordinate, and 0 elsewhere:

\[
u^{n+1}(y; x) = \sum_{j=1}^{2} \left[ x_j \mu_j r_j + y_j \mu_j' r_j + \lambda_j \nu^0_j (y; x) + x_j \mu_j u^n(y; x - e_j) \right. \\
\left. + y_j \mu_j' u^n(y - e_j; x) \right] + \gamma u^n(y; x),
\]  

(1)
where

\[ v_j^n(y; x) = \max \{ u^n(y + e_j; x), u^n(y; x + e_j), u^n(y; x) \} \] and

\[ \gamma = c\mu_2 + \sum_{i=1}^{2} (c_i\mu_i - x_i\mu_i - y_i\mu_i), \]

with \( u^n(s - e_j) = u^n(s) \) if \( s_j = 0 \), and \( u^n(s + e_j) = -\infty \) if \( s_j + 1 > c_j \) for \( j = 1, 2 \) or if \( s_3 + s_4 + 1 > c \). Therefore, if e.g., \( x_1 + x_2 = c \) and \( y_j = c_j \), then \( v_j^n(y; x) = u^n(y; x) \). We define the action \( a^n_j(y; x) \) as the optimal state to move into when a class-\( j \) call arrives at a system in state \((y; x)\) with \( n \) remaining transitions. If a class-\( j \) call arrives, which happens with probability \( \lambda_j \), it is either rejected, that keeps the system in the same state, or it is accepted to either station \( j \), moving the system to state \((y + e_j; x)\), or to station \( 0 \), changing the state to \((y; x + e_j)\). If a class-\( j \) call is completed in station \( j \), with probability \( y_j\mu_j \), the system receives a reward of \( r_j \) and the state changes to \((y - e_j; x)\). If a class-\( j \) call ends in station 0, with probability \( x_j\mu_j \), again a reward of \( r_j \) is gained and the system moves to state \((y; x - e_j)\). The “fictitious” call completions, which occur with probability \( \gamma \) given by equation (3), affect neither the state nor the total reward of the system. Finally, if the system closes down, with probability \( \beta \), the system receives no more reward.

We can observe from the optimality equations (2) that the effect of an additional class-\( j \) call in stations \( j \) and 0, as well as of moving a call from station \( j \) to station 0, is an important quantity. We will see later that the effect of changing a class-\( j \) call to class \( k \) at station 0 is also important. Therefore, we define the following functions:

\[
D^0_j(j0)(y; x) = u^n(y + e_j; x) - u^n(y; x),
\]

\[
D^0_0(j0)(y; x) = u^n(y; x + e_j) - u^n(y; x),
\]

\[
D^n_j(jj)(y; x) = u^n(y + e_j; x) - u^n(y; x + e_j), \] and

\[
D^n_0(12)(y; x) = u^n(y; x + e_1) - u^n(y; x + e_2).
\]

We can interpret the difference \( D^0_j(j0)(y; x) \) as the net benefit of the system due to an additional class-\( j \) call at station \( J, J = 0, j, \) in state \((y; x)\) when there are \( n \) more transitions, and \( D^n_j(jj)(y; x) \) as the net benefit of the system when a class-\( j \) call is moved from station 0 to station \( j \) in state \((y; x + e_j)\), whereas \( D^n_0(12)(y; x) \) is the net benefit of the system when a class-2 call at station 0 is changed to a class-1 call in state \((y; x + e_2)\).
Identifying optimal policies in stations 1 and 2, as well as the existence of preferred class(es) in station 0 require deriving appropriate bounds on $D_n(j, k)(y; x)$. The structure of an optimal policy can be characterized by these bounds only if they are “tight” enough. To obtain such bounds, we need to use the assumptions on service rates, i.e., $\mu_1 \leq \mu_2$ or $\mu_j \leq \mu'_j$ explicitly. Then, coupling becomes the most natural tool. In fact, applying coupling combined with the more standard approach of induction is crucial to derive these bounds.

The coupling procedure can be described generally as follows: We couple two systems so that both will receive the same arrival stream. Coupling of call durations in the two systems is a bit more complicated: If the coupled calls are of the same class and being served in the same station, they depart at the same time, otherwise we use the assumptions that $\mu_1 \leq \mu_2$ or $\mu_j \leq \mu'_j$. If a class-1 call is coupled with a class-2 call, both in station 0, then whenever the coupled class-1 call is completed, the coupled class-2 call also ends with probability 1, due to $\mu_1 \leq \mu_2$. In terms of discrete time, this translates to the following: Both calls leave the system with probability $\mu_1$, and only the class-2 call exits the system with probability $\mu_2 - \mu_1$ leaving the coupled class-1 call in the system. Thus, coupling does not allow a coupled class-1 call to leave the system while the coupled class-2 call is still in progress. Coupling additional class-$j$ calls in stations $j$ and 0 is similar: Both calls leave the system with probability $\mu_j$, and only the call in station $j$ departs with probability $\mu'_j - \mu_j$.

We prove all our results for the objective of maximizing total expected $\beta$-discounted reward for a finite number of transitions, $n$, including the “fictitious” transitions due to the “fictitious” call completions. This provides the powerful tool of induction to prove our results for all $n$: All our proofs are by induction on $n$ combined with coupling. The following initial value function satisfies all the subsequent statements for $n = 0$:

$$u^0(y; x) = \sum_{j=1}^{2} (y_j R'_j + x_j R_j) \quad \forall (y; x) \in S,$$

where $R'_j = \int_0^\infty r_j \mu'_j e^{-\mu'_j t} e^{-\beta t} dt = r_j \mu'_j / (\mu'_j + \beta)$ and $R_j = r_j \mu_j / (\mu_j + \beta)$ are the expected present values of $r_j$ when a class-$j$ call is served in stations $j$ and 0, respectively. In words, $u^0$ collects the rewards of calls in the system at $n = 0$, even if they have not been completed. Of course, this makes no difference in the optimal policy for infinite horizon problems. Now that $u_0$ starts induction for all our statements, all our proofs will only show that the statements are true for $n + 1$, assuming that they hold for $n$. 

September 9, 2003
All the results proven for finite $n$ are true for the limit $n \to \infty$, so the corresponding conclusions are valid when total expected $\beta$-discounted reward over an infinite horizon is maximized. Moreover, since the state space and the action space in each state are finite and the results hold even for $\beta = 0$, we have the same conclusions for maximizing the long-run average reward.

IV. OPTIMAL ADMISSION POLICY AT STATION 1 AND 2

This section specifies the optimal admission policy for stations 1 and 2. Lemma 1 establishes certain bounds on $D^n_j(jk)$, with very intuitive implications on optimal admission. Hence, although their formal proofs require some work, we omit them and refer to [11].

Lemma 1: (i) For all $(y + e_j; x) \in S$: $0 \leq D^n_j(0)(y; x) \leq R'_j$ for all $n \geq 0$ and for $j = 1, 2$.
(ii) For all $(y; x + e_1) \in S$: $D^n_0(j0)(y; x) \leq R_j$ for all $n \geq 0$ and for $j = 1, 2$.
(iii) For all $(y + e_j; x + e_1) \in S$: $0 \leq D^n_{j0}(jj)(y; x)$ for all $n \geq 0$ and for $j = 1, 2$.
(iv) For all $(y; x + e_1) \in S$: $D^n_0(12)(y; x) \leq R_1 - R_2$ for all $n \geq 0$.

Parts (ii) and (iv) provide upper bounds on the benefit of the system due to an additional class-$j$ call, and due to converting a class-2 call to class 1 in station 0, which are used in proving subsequent statements, see [11]. Parts (i) and (iii) of Lemma 1 imply that it is always better to accept a class-$j$ call to station $j$ rather than rejecting it, and rather than accepting it to station 0, respectively. This specifies the optimal admission rules for stations 1 and 2 due to optimality equations (2):

Theorem 1: If $y_j < c_j$, it is optimal to admit a class-$j$ call to station $j$, $j = 1, 2$.

V. EXISTENCE OF A PREFERRED CLASS IN STATION 0

A consequence of Theorem 1 is that admission of calls to station 0 should be considered only after the dedicated station of class $j$ becomes completely full. Then station 0 behaves like a stochastic knapsack receiving an arrival stream equivalent to the overflow process of stations 1 and 2. Örmeci et al. [12] and Savin et al. [14] define a customer class as “preferred”, in the sense that its customers are always admitted to the system if there are free servers, regardless of the congestion level. We adopt their terminology for station 0 so that class $j$ is preferred if it is always optimal to admit class-$j$ calls to station 0 whenever station 0 has at least one available server while $y_j = c_j$. In this section, we derive sufficient conditions for each class to be preferred.
From optimality equations (2), it is clear that if $D_0^j(j0)(y;x) \geq 0$ for all $(y;x+e_1) \in S$, then it is always better to serve a class-$j$ call at station $0$ rather than rejecting it. Our next lemma presents sufficient conditions for $D_0^j(j0)(y;x) \geq 0$ for all $(y;x+e_1) \in S$ for $j = 1, 2$:

**Lemma 2:** (i) If $\lambda_2 r_2 \mu_2 \leq (\lambda_2 + \mu_2 + \beta) r_1 \mu_1$, then for all $(y;x+e_1) \in S$ and for all $n$: $D_0^n(10)(y;x) \geq 0$ and $D_0^n(12)(y;x) \geq (r_1 \mu_1 - r_2 \mu_2) / (\mu_2 + \beta)$.

(ii) If $\lambda_1 R_1 \leq (\lambda_1 + \mu_2 + \beta) R_2$, then for all $(y;x+e_1) \in S$: $0 \leq D_0^n(20)(y;x)$ for all $n \geq 0$.

**Proof:** Both proofs are similar: We prove part (i), and refer to [11] for part (ii).

We first consider the statement $D_0^n(10)(y;x) \geq 0$. Assume that system A is in state $(y;x+e_1)$ and system B is in $(y;x)$ in period $n+1$. If an arrival occurs while $y_j < c_j$, both systems accept the new call to station $j$, keeping the difference between the two systems the same. So consider an arrival when $y_j = c_j$: System A rejects all calls and system B follows the optimal policy. If upon an arrival system B also rejects the incoming call, both systems remain in their current states. Acceptance of a class-1 call to system B couples the systems. If system B admits a class-2 call, then the systems move to states $(y;x+e_1)$ and $(y;x+e_2)$. With the departure of the additional class-1 call in system A, the systems couple with an additional reward of $r_1$ for system A, whereas all other service completions keep the difference between the two systems the same. Then:

$$D_0^{n+1}(10)(y;x) \geq \lambda_1 \min \left\{ \min_{s+e_1 \in S} \left\{ D_0^n(10)(s) \right\}, 0 \right\}$$

$$+ \lambda_2 \min \left\{ \min_{s+e_1 \in S} \left\{ D_0^n(10)(s) \right\}, D_0^n(12)(y;x) \right\}$$

$$+ \mu_1 r_1 + \left( \sum_{j=1,2} c_j \mu_j + (c-1) \mu_2 \right) \min_{s+e_1 \in S} \left\{ D_0^n(10)(s) \right\}$$

$$\geq \lambda_2 \min \left\{ 0, \frac{r_1 \mu_1 - r_2 \mu_2}{\mu_2 + \beta} + \lambda_1 \right\}$$

where the first inequality is due to coupling, and the second follows from the induction hypotheses. If $r_1 \mu_1 \geq r_2 \mu_2$, the statement is proven; otherwise:

$$D_0^{n+1}(10)(y;x) \geq r_1 \mu_1 \frac{\lambda_2 + \mu_2 + \beta}{\mu_2 + \beta} - \frac{\lambda_2 r_2 \mu_2}{\mu_2 + \beta} \geq 0,$$

where the last inequality is due to the assumption of the theorem.

Now consider the statement for $D_0^n(12)(y;x)$. Let system A be in state $(y;x+e_1)$ and system B in $(y;x+e_2)$ in period $n+1$. System B takes the optimal actions and system A imitates all
the actions of system B. We couple the additional class-2 call in system B with the additional class-1 call in system A, as well as all other service and interarrival times. Then, with probability \( \mu_1 \), the systems couple with a difference in reward, \( r_1 - r_2 \). With probability \( \mu_2 - \mu_1 \), the systems move to two different states, \( (y; x + e_1) \) and \( (y; x) \) with a difference of \(-r_2\). Whenever there is any other transition, both systems continue to have their additional calls. Thus:

\[
D_{0}^{n+1}(12)(y; x) \geq \mu_1(r_1 - r_2) + (\mu_2 - \mu_1)(D_{0}^{n}(10)(y; x) - r_2) + \left( \sum_{j=1,2} (\lambda_j + c_j \mu_j') + (c-1)\mu_2 \right) \min_{s+e_3 \in S} \{D_{0}^{n}(12)(s)\}
\]

\[
\geq r_1\mu_1 - r_2\mu_2 + (1 - \mu_2 - \beta) \frac{r_1\mu_1 - r_2\mu_2}{\mu_2 + \beta} = \frac{r_1\mu_1 - r_2\mu_2}{\mu_2 + \beta}
\]

where the first inequality is due to the coupling and the second follows by uniformization and the induction hypotheses. This proves part (i) of the lemma.

This lemma shows that under the specified conditions it is always better to accept a class-\( j \) call to station 0 rather than rejecting it. We know from Theorem 1 that station 0 should admit only the overflow of stations 1 and 2. Hence, we can conclude the following:

**Theorem 2:** (i) If \( y_1 = c_1 \) and \( \lambda_2 r_2 \mu_2 \leq (\lambda_2 + \mu_2 + \beta)r_1 \mu_1 \), then it is optimal to accept a class-1 call to station 0 whenever it has available server(s), i.e., class 1 is preferred.

(ii) If \( y_2 = c_2 \) and \( \lambda_1 R_1 \leq (\lambda_1 + \mu_2 + \beta)R_2 \), then it is optimal to accept a class-2 call to station 0 whenever it has available server(s), i.e., class 2 is preferred.

Under the conditions specified by Theorem 2, it is optimal to accept class-\( j \) calls to station 0 regardless of the precise state of station 0 or of station \( k, k \neq j \), i.e., it is enough to have at least one available server at station 0 while station \( j \) is completely busy. In fact, the conditions on the parameters of the system are exactly the same as those in [12], where the stochastic knapsack receives two Poisson arrivals with rates \( \lambda_1 \) and \( \lambda_2 \). At first sight, this is surprising since we would expect a dependence on the total service capacity of station \( j \) determined by \( c_j \) and \( \mu_j \).

However, Lemma 2 is a result of the trade-off between losing a call and occupying a server of station 0 given the total potential arrival rate at station 0. Hence, Lemma 2 is not concerned with when a call arrives at station 0. Therefore, the effect of station \( j \) on the admission in station 0 appears only through the current state of the system with the condition \( y_j = c_j \). These issues are discussed further in section VII.

September 9, 2003 DRAFT
VI. AN OPTIMAL THRESHOLD POLICY FOR STATION 0

We expect that it should be more difficult to accept calls to station 0 when there are many calls already in the system. As station 0 has more calls, the resources shared by both classes decrease, which in turn decreases the marginal benefit of the system from an additional call in station 0. Part (i) of Lemma 3 shows that the benefit of an additional class-\(j\) call in station 0 is decreasing in \(x_k\), \(k \neq j\). We would also expect that this benefit decreases in \(x_j\), which we could not prove due to disturbing effects at the boundary. We discuss this issue further in section VII via numerical examples. On the other hand, when station \(k\) has more calls, the probability of overflow in station \(k\) increases. Hence, as \(y_k\) increases, the system becomes more reluctant to admit class-\(j\) calls to station 0 since the resources in station 0 may be needed by future class-\(k\) calls soon. Part (ii) of Lemma 3 proves that the benefit of an additional class-\(j\) call in station 0 decreases in \(y_k\).

Lemma 3: For all \(n \geq 0\) and for \(j = 1, 2\) and \(k \neq j\):

(i) For all \((y; x + e_k + e_j) \in S\):

\[
    u^n(y; x + e_k + e_j) - u^n(y; x + e_k) \leq u^n(y; x + e_j) - u^n(y; x).
\]

(ii) For all \((y + e_k; x + e_j) \in S\):

\[
    u^n(y + e_k; x + e_j) - u^n(y + e_k) \leq u^n(y; x + e_j) - u^n(y; x).
\]

Proof: We prove part (ii) only, and refer to [11] for part (i).

First, note that (6) implies and is in fact equivalent to:

\[
    u^n(y + b_k e_k; x + b_j e_j) - u^n(y + b_k e_k; x) \leq u^n(y; x + b_j e_j) - u^n(y; x).
\]

We want to show that \(v^n_i\)'s also satisfy this monotonicity. Hence we define:

\[
    \delta^n_i = v^n_i(y + e_k; x + e_j) - v^n_i(y + e_k; x) - v^n_i(y; x + e_j) + v^n_i(y; x).
\]

Let systems A, B, C and D be systems starting from states \((y + e_k; x + e_j)\), \((y + e_k; x)\), \((y; x + e_j)\), and \((y; x)\), respectively, in period \(n + 1\).

If a class-\(k\) call arrives when \(y_k + 1 < c_k\), then all four systems accept the call to station \(k\), satisfying the inequality trivially by the induction hypothesis. Now assume \(y_k + 1 = c_k\): In this case system C and D will follow their optimal policy, thus accepting the incoming call to station
k, whereas system A and B will accept \(a_A\) class-\(k\) calls to station 0 with \(a_A\) being the optimal number of calls that system A accepts to station 0. Then:

\[
\delta^n_k \leq u^n(y + e_k; x + a_A e_k + e_j) - u^n(y + e_k; x + a_A e_k) - u^n(y + e_k; x + e_j) + u^n(y + e_k; x) \leq 0,
\]

where the first inequality is true since \(v^n_k(y + e_k; x) \geq u^n(y + e_k; x + a_A e_k)\) by definition of \(v^n_k\)'s. If \(a_A = 0\), then the expression becomes 0, otherwise, i.e., if \(a_A = 1\), the inequality becomes inequality (5) with \((y; x) = (y + e_k; x)\) and so it is satisfied by part (i).

If a class-\(j\) arrival occurs when \(y_j < c_j\), then all four systems accept the call to station \(j\) so that \(\delta^n_j \leq 0\) by the induction hypothesis. Hence, consider the states with \(y_j = c_j\): Systems A and D take the optimal actions, where \(a_A\) and \(a_D\) are the number of calls accepted to station 0 in systems A and D, respectively. Then, systems B and C accept \(a_D\) and \(a_A\) calls to station 0, respectively, which is always feasible as systems B and C have the same number of calls in station 0 with systems D and A, respectively. We have:

\[
\delta^n_j \leq u^n(y + e_k; x + (1 + a_A) e_j) - u^n(y + e_k; x + a_D e_j) - u^n(y; x + (1 + a_A) e_j) + u^n(y; x + a_D e_j) \leq 0,
\]

due to definition of \(v^n_j\)'s, and since the second inequality is a special case of inequality (7).

We couple all these systems so that, except for the additional calls, they all behave in the same way. Moreover, we couple the additional class-\(k\) calls in systems A and B, and the additional class-\(j\) calls in systems A and C. Before considering \(u^{n+1}\), we define:

\[
\Delta^n(y; x) = u^n(y + e_k; x + e_j) - u^n(y + e_k; x) - u^n(y; x + e_j) + u^n(y; x)
\]

Then:

\[
\Delta^{n+1}(y; x) = \sum_{l=j,k} \left( \lambda_l \delta^n_l + x_l \mu_l \Delta^n(y; x - e_i) + y_l \mu'_l \Delta^n(y - e_i; x) \right) + y^l \Delta^n(y; x)
\]

\[
+ \mu_j [u^n(y + e_k; x) - u^n(y + e_k; x) - u^n(y; x) + u^n(y; x)]
\]

\[
+ \mu'_k [u^n(y; x + e_j) - u^n(y; x) - u^n(y; x + e_j) + u^n(y; x)] \leq 0,
\]

with \(y^l = c \mu_2 - \mu_j - \mu'_k + \sum_{i=1}^2 (c_i \mu'_i - x_i \mu_i - y_i \mu'_l)\). We have shown \(\delta^n_j \leq 0\), while the terms (2-4) are also non-positive by the induction hypothesis. Finally, the fifth and the sixth terms are 0. Thus, the value functions, \(u^n\), satisfy inequality (6) for all \(n\).
Lemma 3 considers monotonicity properties of the value functions for the whole system. Part (i) shows submodularity for the shared station, while the states of the dedicated stations are taken into account explicitly. This guarantees the existence of optimal thresholds even in the presence of stations 1 and 2. Moreover, part (ii) establishes the monotonicity of the thresholds in the number of calls in station $k$. Hence, the arrivals to the shared station cannot be viewed simply as an overflow process when monotonicity properties are considered. Theorem 3 states these conclusions:

Theorem 3: For $j = 1, 2$, there exist numbers $\left\{l^n_j(y_k;0),...,l^n_j(y_k;c - 1)\right\}$ such that it is optimal to accept a class-$j$ call to station 0 if $y_j = c_j$ and $x_k < l^n_j(y_k;x_j)$, and to reject it otherwise, where $k \neq j$. Moreover, $l^n_j(y_k;m) \geq l^n_j(y_k+1;m)$.

VII. ILLUSTRATION AND DISCUSSION OF RESULTS VIA NUMERICAL EXAMPLES

In our first example, stations 1 and 2 both have 4 servers, while station 0 has 6. The other parameters are: $\lambda_1 = 25, \lambda_2 = 125, r_1 = 1.8, r_2 = 0.255, \mu_1 = 0.75, \mu_2 = 6, \mu'_1 = 1$, and $\mu'_2 = 8$. For this specific example, our results, cannot specify a preferred class, in fact they do not even guarantee the existence of one. However, as in all other examples we have generated, the system has a preferred class: Class 2 is preferred, so its thresholds are trivial. Figure 2 presents optimal thresholds for class 1: $l_1(y_2,x_1)$ is decreasing in $y_2$ for fixed $x_1$ as a consequence of Theorem 3. Moreover, $l_1(y_2,x_1)$ decreases in $x_1$ for fixed $y_2$, in accordance with our intuition about the increasing reluctance of the system to admit class-$j$ calls as a reaction to increasing number of class-$j$ calls already in station 0. This intuition could be established by concavity of the value functions, which is shown only under restrictive conditions even in the usual stochastic knapsacks (see [12] and [3]).

Next, we investigate the issue of preferred class: We increase the number of operators available to the dedicated stations to 28, while keeping the rest of the parameters the same. When $c_1 = 12$ and $c_2 = 16$, class 1 is preferred: Station 0 rejects class-2 calls in a few states when both dedicated stations are completely occupied. If $c_1 = 17$ and $c_2 = 11$, then all calls are admitted to station 0. Finally, when $c_1 = 22$ and $c_2 = 6$, class 2 is preferred as station 0 denies service to class 1 in a few states, when all servers in both stations 1 and 2 are busy. These examples show that preferred class depends on the capacities of the dedicated stations, since they determine the effective arrival rates to station 0. For a detailed discussion on the relation between the arrival
rates and preferred class(es), see [12].

**VIII. CONCLUSION**

In this paper, we analyze the dynamic admission control of a loss system with one shared and two dedicated stations. The main purpose of this analysis is to provide insight on the design and control of large call centers receiving two classes of calls. However, this kind of models is also encountered in certain telecommunications and manufacturing systems, where the assumption of no waiting room becomes critical. Recent trends in these systems, such as just-in-time manufacturing and wider use of synchronous services on the internet, e.g., real-time video or audio, demand for loss models with admission control (see for example [9]). Our model is suitable for systems, which utilize admission restrictions rather than scheduling and routing.
in order to control the amount of work in the system. Still, our future research plan includes considering finite buffers as well as more than two call classes that will widen applicability of our model.

REFERENCES


