A Feedback Control Model with Negligible Switching Costs for a Retail Facility

D. Arivudainambi and V. Poongothai

Department of Mathematics
Anna University
Chennai-600 025, India
arivu@annauniv.edu, rvgothai@gmail.com

Abstract

We consider a retail service facility with cross trained servers who can perform operations in both front room and back room. The front room server serves the customers who arrive to the facility and the back room servers a job which is generated by the front room. Two essential models for service facility are discussed in this paper. For the fixed number of cross-trained servers, the objective is to find an optimal policy so that the expected number of customer in the queue is minimized subject to a back room constraint is discussed in the first model. In the second model, the goal is to minimize the number of servers subject to both front room and back room constraints. After completing the front room service if the customer is unsatisfied with the service, he may rejoin for service again until the service is completed successfully. The paper includes an analysis of the model and several effective algorithms were designed to solve the models. The effects of various parameters on the performance measures are also analyzed numerically.

Keywords: Cross trained, Front room and back room operations, Optimal control, Optimization

1 Introduction

We consider a retail facility with cross trained servers who can perform operations in front room and back room. The server directly serves the customer is known as front room service and the server serves a job generated in the front room is known as back room service. Retail service facilities of this kind include banks, insurance agencies, retail stores, post offices, etc. For examples in banks, service to the customers in the front room are granting a loan, open a new account, etc and updating the customer information, performing credit
checks, etc are work done in the back room. The management can improve the quality of service by switching the servers from front room to back room or vice versa. After completion of the front room service, if the customer is unsatisfied with the service, he may rejoin the service again to the front room until the service is completed successfully.

Our goal is to determine an optimal switching policy to the servers so that the expected customer waiting time in the queue is minimized subject to the back room constraint. A switching policy can be defined as the number of servers assigned to the front room depending on the number of customer waiting in the queue and the amount of work in the back room.

Two models are analyzed briefly in this paper. Queue control problem is considered in the first model. Given a fixed number of cross-trained servers, the objective is to minimize the expected customer waiting time in the front room subject to back room constraint by switching the servers. Queue design and queue control problem is discussed in the second model. Queue design model is to determine the optimal value of parameters which will become a fixed characteristics of the queue. Our goal is to minimize the number of servers subject to back room and front room constraints. We assume there is no loss of productivity due to switching of servers from front room to back room and vice versa.

Berman and Sapna [5] considered a retail facility with the front room and the back room operations in which the amount of back room work is correlated with the amount of work in the front room. Moder and Phillips [9] dealt with controlling the number of servers based on the queue length by increasing a server one at a time when the queue length becomes greater. In Yadin and Naor [14], a single server queue with controllable service rate is analyzed, and the expected system size is obtained. Wang [13] considered the state of the art of the queueing theory with a variable number of servers in which a loss of productivity occurs when a servers returns to the back room. Loss of productivity due to switching of servers from front room to back room and a non deterministic policy with cross trained servers is taken into account by Berman and Ianovsky [3].

The control of arrival and service rates in a single server queue without switching costs is studied in Serfozo [10]. The work of Serfozo and Lu [11] is interesting to characterize the optimal policy in case of more service levels. Terekhov and Beck [12] considered the problem of finding the optimal combination of cross-trained and specialized servers. Queueing control models are discussed in the work of Berman and Larson [4] by considering three dimensional state variables with two different threshold levels. A model of the beginning-of-shift allocation problem is discussed in Campbell and Diaby [7]. Campbell used the model to investigate the value of cross training on the development and evaluation of an effective heuristic solution algorithm.
Agnihothri et al. [2] considered a field service system with two types of jobs and a fixed number of servers to minimize the sum of the average service costs and the customer delay costs per unit time. Kitaev and Serfozo [8] studied an $M/M/1$ queue with switching costs, where the switching costs are based on the choice of arrival and service rates. Unsatisfied customers demand is more important in any service facility which is not considered in all the above papers. To fill this gap, we considered a service facility which can handle unsatisfied customers demand too.

The rest of the paper is organized as follows. A brief mathematical description of queue design and queue control is given in the section 2 and is analyzed in section 3. The algorithms for solving queueing design and control problem is discussed in section 4. To illustrate the effect of the parameters on the main performance measures, numerical analysis is done in section 5.

## 2 Mathematical Model

Let $R$ be the maximum number of customers allowed in the front room and $N$ be the number of servers in the system. Customers arrive according to a Poisson process with rate $\lambda$. The service time in the front room is assumed to be exponentially distributed with service rate $\mu$. After completion of front room service if the customer is not satisfied with the service the customer may immediately join for second optional service with probability $\theta$ or may leave the system with probability $1 - \theta$.

If there are already $R$ customers in the front room, new arrivals will be blocked from joining the queue, since the capacity of the system is finite. Let $B_s$ denotes the expected number of servers require to complete the work in the back room. We assume that the switching time and the switching cost are negligible.

Let $P(s)$ be the steady-state probability of having $s$ customers in the system for $s = k_0, k_0 + 1, \ldots, k_N$, where $k_0$ is a non-negative integer. A policy described in terms of quantities $k_r, r = 0, 1, \ldots, N$, whenever there are between $k_{r-1} + 1$ and $k_r$ customers in the queue, there should be $r$ servers in the front room. The number of customers waiting in the queue and the amount of work in the back room depends on the number of servers assigned to the front room which defines a switching policy.

We assume that $|k_r - k_{r-1}| \geq 1$ and $k_N = R$ (i.e.) all $N$ servers are employed in the front room till the system reaches its capacity $R$. The set of balance equations $(E_1)$ to $(E_N)$ with normalization condition $\sum_{s=k_0}^{k_N} P(s) = 1$ determines the steady state probabilities uniquely.

The following balance equations establishes the relationship between $P(s)$ and
\( P(s + 1) \) for \( s = k_0, k_0 + 1, \ldots, k_N. \)

\[ (E_1) \quad P(s) \lambda = P(s + 1)(1 - \theta)\mu + P(s)\theta\mu, \quad s = k_0, k_0 + 1, \ldots, k_1 - 1 \]

\[ (E_2) \quad P(s) \lambda = P(s + 1)(1 - \theta)2\mu + P(s)\theta2\mu, \quad s = k_1, k_1 + 1, \ldots, k_2 - 1 \]

\[ \vdots \]

\[ (E_r) \quad P(s) \lambda = P(s + 1)(1 - \theta)r\mu + P(s)\theta r\mu, \quad s = k_r - 1, \ldots, k_r - 1 \]

\[ \vdots \]

\[ (E_N) \quad P(s) \lambda = P(s + 1)(1 - \theta)N\mu + P(s)\theta N\mu, \quad s = k_N - 1 \ldots, k_N - 1 \]

In steady state, the average flow from state \( s + 1 \) to the state \( s \) is equal. The mean flow from state \( s \) to \( s + 1 \) is \( P(s) \lambda \) and the mean flow from state \( s + 1 \) to \( s \) is \( P(s + 1)\mu \) for \( r = 1, 2, \ldots, N \).

From balance equation \((E_1)\), we get

\[
P(s + 1) = \left( \frac{\lambda - \theta \mu}{(1 - \theta) \mu} \right) P(s), \quad s = k_0, k_0 + 1, \ldots, k_1 - 1
\]

The steady-state probability of having \( s \) customers in the front room can be obtained from \((E_1)\) by substituting \( s = k_0, s = k_0 + 1, \ldots, s = k_N \) and we get

\[
P(s) = \left( \frac{\lambda - \theta \mu}{(1 - \theta) \mu} \right)^{s - k_0} P(k_0), \quad s = k_0, k_0 + 1, \ldots, k_1
\]

Substituting \( s = k_1 \) in the above, we get

\[
P(k_1) = \left( \frac{\lambda - \theta \mu}{(1 - \theta) \mu} \right)^{k_1 - k_0} P(k_0)
\]

The steady-state probability can be obtained from \((E_2)\) by substituting \( s = k_1, s = k_1 + 1, \ldots, s = k_N \) and combining all equations, we get

\[
P(s) = \left( \frac{\lambda - 2\theta \mu}{2(1 - \theta) \mu} \right)^{s - k_1} P(k_1)
\]

Substituting the value of \( P(k_1) \) in \((E_2)\), we get the steady state probability as

\[
P(s) = \left( \frac{\lambda - 2\theta \mu}{2(1 - \theta) \mu} \right)^{s - k_1} P(k_1) = \left( \frac{\lambda - 2\theta \mu}{2(1 - \theta) \mu} \right)^{s - k_1} \left( \frac{\lambda - \theta \mu}{(1 - \theta) \mu} \right)^{k_1 - k_0} P(k_0)
\]

The steady state probability can be obtained

\[
P(s) = \left( \frac{\lambda - N\theta \mu}{N(1 - \theta) \mu} \right)^{s - k_N} P(k_N)
\]

\[
P(s) = \left( \frac{\lambda - r\theta \mu}{r(1 - \theta) \mu} \right)^{s - k_{r-1}} \left( \frac{\lambda - (r - 1)\theta \mu}{(r - 1)(1 - \theta) \mu} \right)^{k_{r-1} - k_{r-2}} \ldots \left( \frac{\lambda - \theta \mu}{(1 - \theta) \mu} \right)^{k_1 - k_0} P(k_0)
\]

\[
P(s) = \left( \frac{\lambda - \mu \theta}{\mu(1 - \theta)} \right)^{s - k_0} \left( \frac{1}{r} \right)^{s - k_{r-1}} X_r, \quad s = k_{r-1} + 1, \ldots, k_r
\]
The steady state probability can be expressed as,

\[ P(s) = B_s P(k_0) \]

where \( B_s = \left\{ \begin{array}{ll} 1, & s = k_0 \\ \left( \frac{\lambda - \mu \theta}{\mu (1 - \theta)} \right)^{s-k_0} \left( \frac{1}{r} \right)^{s-k_{r-1}} X_r, & k_r, r = 1, 2, \ldots N \end{array} \right. \]

\[ X_r = \prod_{h=1}^{r-1} \left( \frac{1}{h} \right)^{k_h-k_{h-1}} (X_1 = 1, r = 1, 2, \ldots N + 1) \]

Also, \( P(k_0) \sum_{s=k_0}^{k_N} B_s = 1 \).

The expected number of servers in the front room is obtained as,

\[ F = \sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} r P(s) = \frac{\sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} r B_s}{\sum_{s=k_0}^{k_N} B_s} \]

The expected number of servers in the back room is expressed as \( B = N - F \).

The expected number of customers in the front room is given by

\[ L = \sum_{s=k_0}^{k_N} s P(s) = P(k_0)\left[ k_0 + \sum_{s=k_0+1}^{k_N} sB_s \right] \]

The expected waiting time in the queue \( W_q \) is obtained using Little’s formula

\[ W_q = \frac{L}{\lambda (1 - P(k_N))} - \frac{1}{\mu} = \frac{k_0 + \sum_{s=k_0+1}^{k_N} sB_s}{\sum_{s=k_0}^{k_N-1} B_s} - \frac{1}{\mu} \]

In model \( M_1 \), for the number of servers \( N \), we find the optimal switching points \( k_0, k_1, \ldots k_{N-1} \) and \( k_N = R \), so as to minimize the expected waiting time of customers in the queue subject to the back room constraint. Let \( \mathcal{R} = \{ T; T = \{ k_0, k_1, \ldots k_N, R \} \text{ be the family of all possible switching policies,} \)
where \( k_r \) are integers and \( |k_r - k_{r-1}| \geq 1, k_0 \geq 0, k_{N-1} < s \} \).

### 2.1 Model \( M_1 \): Optimal Control of Queues

The optimal control of a queueing system is to determine arrival or service rates to optimize some objective function. Problems dealing with queue control are dynamic in nature. The objective is to find the optimal policy to minimize the length of the queue. Queue control optimization problems are normally solved via dynamic programming, linear programming, etc. In model \( M_1 \), given a fixed number of servers \( N \), subject to a back room constraint,

\[ \min_{T \in \mathcal{R}} W_q \]

subject to \( B \geq B_s \) \hspace{1cm} (M_1)
2.2 Model $M_2$: Optimal Design and Queue Control

The optimal design of a queueing system is to determine the optimal value of parameters, such as optimal mean service rate or optimal number of servers. Problems dealing with queue design are static in nature. The goal is to find the smallest number of cross trained servers. In general, queueing design optimization problems can be solved using differential calculus and linear, integer or non-linear programming, algorithms. In model $M_2$, we find the minimum number of servers $N$, subject to the same constraint as in $(M_1)$ with an additional constraint to ensure that $W_q$ is bounded from above as well.

$$\min_{T \in \mathbb{R}} N$$

subject to $W_q \leq W_u$

$B \geq B_s$ \hspace{1cm} $(M_2)$

It is quite stronger in case of negligible switching cost. The effective time average number of servers in the back room decreases with respect to loss efficiency for the two models. The optimal switching points will increase for large lower bound $B_s$ and result in a smaller $F$ but a higher $W_q$. In model $(M_2)$, we get a higher $N$ value at optimal (i.e) more servers are needed.

3 Analysis of the Model

Let us define the function $f \in \{0, 1, \cdots, N-1\}$ and two policies $T = (k_0, k_1, \cdots, k_N)$, $V = (k'_0, k'_1, \cdots, k'_N)$, $T, V \in \mathbb{R}$, where $|k_r - k_{r-1}| \geq 1, |k_f - k_{f-1}| \geq 2$, $k'_r = k_r$ for $r \neq f$ and $k'_f = k_f - 1$. Let $F(T), B(T)$ and $W_q(T)$ be the expected number of servers in front room, back room and waiting time in the queue under the policy $T$. Similarly $F(V), B(V)$ and $W_q(V)$ be the expected number of servers in front room, back room and waiting time in the queue under the policy $V$. Let $P(s)$ and $P'(s)$ be the steady state probabilities corresponding to $T$ and $V$ respectively and $B_s = \frac{P(s)}{P(k_0)}$ and $B'_s = \frac{P'(s)}{P(k_0)}$ where $B_s$ and $B'_s$ be the expected number of servers required to complete the work in the back room with policies $T$ and $V$ respectively.

3.1 Theorem

For the two policies $T$ and $V$,

a) $W_q(T) \geq W_q(V)$; b) $F(T) \geq F(V)$; c) $B(T) \geq B(V)$;

Proof

Let $P(s)$ and $P'(s)$ be the steady state probabilities corresponding to policies $T$ and $V$ respectively. The expected number of servers in the front room and
From (6) and (8) the difference between the expected waiting time in the queue with respect to policies $T$ and $V$ can be expressed as follows:

$$ F(T) = \frac{\sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} rB_s}{\sum_{s=k_0}^{k_N} B_s} $$

(5)

$$ W_q(T) = \frac{k_0 + \sum_{s=k_0+1}^{k_N} sB_s}{\sum_{s=k_0}^{k_N} B_s} - \frac{1}{\mu} $$

(6)

Therefore $N/D > 0$ and thus $W_q(T) - W_q(V) > 0$. 

The expected number of servers require to complete the work in the back room with respect to policy $T$ and $V$ can be expressed as follows:

$$ B'_s = \left\{ \begin{array}{ll} B_s, s = k_0, k_0 + 1, \cdots k_f - 1 \\ \frac{f}{f+1} B_s, s = k_f, k_f + 1, \cdots k_N \end{array} \right. $$

(7)

(a) Using (7) in the equation (4), the expected waiting time in the queue with policy $V$ can be expressed as follows:

$$ W_q(V) = \frac{k_0 + \sum_{s=k_0+1}^{k_N} sB_s + \sum_{s=k_0+1}^{k_f} \frac{f}{f+1} sB_s}{\lambda \left( \sum_{s=k_0}^{k_f} B_s + \sum_{s=k_f}^{k_N} \frac{f}{f+1} B_s \right)} - \frac{1}{\mu} $$

(8)

From (6) and (8) the difference between the expected waiting time in the queue with policies $T$ and $V$ is expressed as

$$ W_q(T) - W_q(V) = \frac{k_0 + \sum_{s=k_0+1}^{k_N} sB_s}{\sum_{s=k_0}^{k_N} B_s} - \frac{k_0 + \sum_{s=k_0+1}^{k_N} sB_s + \sum_{s=k_f}^{k_N} \frac{f}{f+1} sB_s}{\lambda \left( \sum_{s=k_0}^{k_f} B_s + \sum_{s=k_f}^{k_N} \frac{f}{f+1} B_s \right)} $$

The expected waiting time can be expressed as a ratio $N/D$ where

$$ D = \lambda \left( \sum_{s=k_0}^{k_{f-1}} B_s \right) \left( \sum_{s=k_0}^{k_{f-1}} B_s + \sum_{s=k_f}^{k_{N-1}} \frac{f}{f+1} B_s \right) > 0 $$

$$ N = \sum_{b=k_0}^{k_N} \sum_{s=k_0}^{k_f} bB_bB_s + \sum_{b=k_0}^{k_N} \sum_{s=k_f}^{k_f} \frac{f}{f+1} bB_bB_s - \sum_{b=k_0}^{k_N} \sum_{s=k_0}^{k_f} bB_bB_s $$

$$ - \sum_{b=k_f}^{k_N} \sum_{s=k_0}^{k_{f-1}} \frac{f}{f+1} bB_bB_s $$

$$ N = \frac{1}{f+1} \left( \sum_{b=k_f}^{k_{N-1}} \sum_{s=k_0}^{k_f} (b-s)B_bB_s + \sum_{b=k_0}^{k_N} kNB_bB_s \right) > 0 $$

Therefore $N/D > 0$ and thus $W_q(T) - W_q(V) > 0$. 

(b) Using (7) in the equation (3) the expected number of servers in the front room with policy $V$ can be expressed as follows:

$$F(V) = \frac{\sum_{r=1}^{f} \sum_{s=k_{r-1}+1}^{k_r} rB_s + \sum_{r=f+1}^{N} \sum_{s=k_{r-1}+1}^{k_r} (\frac{1}{f+1}) rB_s}{\sum_{s=k_0}^{k_{f-1}} B_s + \frac{f}{f+1} \sum_{s=k_f}^{k_N} B_s}$$  \hspace{1cm} (9)

From (5) and (9) the difference between the expected number of servers in the front room with policies $T$ and $V$ is expressed as

$$F(V) - F(T) = \frac{\sum_{r=1}^{f} \sum_{s=k_{r-1}+1}^{k_r} rB_s + \sum_{r=f+1}^{N} \sum_{s=k_{r-1}+1}^{k_r} (\frac{1}{f+1}) rB_s}{\sum_{s=k_0}^{k_{f-1}} B_s + \frac{f}{f+1} \sum_{s=k_f}^{k_N} B_s}$$

The difference between $F(V)$ and $F(T)$ can be expressed as a fraction $G/E$

where $G = \sum_{r=1}^{f} \sum_{s=k_{r-1}+1}^{k_r} \sum_{g=k_0}^{k_N} rB_s B_g + \sum_{r=f+1}^{N} \sum_{s=k_{r-1}+1}^{k_r} \sum_{g=k_0}^{k_N} (\frac{f}{f+1}) rB_s B_g - \sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} \sum_{g=k_0}^{k_N} rB_s B_g - \sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} \sum_{g=k_f}^{k_N} (\frac{f}{f+1}) rB_s B_g$

$$E = \left( \sum_{s=k_0}^{k_N} B_s \right) \left( \sum_{s=k_f}^{k_N} B_s + \frac{f}{f+1} \sum_{s=k_f}^{k_N} B_s \right) > 0$$

$$G = \frac{1}{f+1} \left[ \sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} rB_s B_{k_f} - \sum_{r=f+1}^{N} \sum_{s=k_{r-1}+1}^{k_r} rB_s \right.$$  

$$\left. + \sum_{r=1}^{f} \sum_{s=k_{r-1}+1}^{k_r} \sum_{m=f+1}^{N} \sum_{l=k_{m-1}+1}^{k_m} (r-m)B_s B_l \right]$$

Using (2), $G$ can be expressed as

$$G = \frac{1}{f+1} \left[ \sum_{r=1}^{N} \sum_{s=k_{r-1}+1}^{k_r} \left( \frac{1}{r} \right) s^{-k_{r-1}-1} \left( \frac{\lambda - \mu \theta}{\mu (1-\theta)} \right)^{s+k_{f-2}k_0} \left( \frac{1}{f} \right) k_f^{-k_{f-1}} X_r X_f \right.$$  

$$\left. + \sum_{r=1}^{f} \sum_{s=k_{r-1}+1}^{k_r} \sum_{m=f+1}^{N} \sum_{l=k_{m-1}+1}^{k_m} (r-m) \left( \frac{1}{r} \right) s^{-k_{r-1}-1} \left( \frac{1}{m} \right) l^{-k_{m-1}} X_r X_l \right.$$  

$$\times \left( \frac{\lambda - \mu \theta}{\mu (1-\theta)} \right)^{s+l-2k_0} - \sum_{r=f+1}^{N} \sum_{s=k_{r-1}+1}^{k_r} \left( \frac{1}{r} \right) s^{-k_{r-1}-1} \left( \frac{\lambda - \mu \theta}{\mu (1-\theta)} \right)^{s-k_0} X_r \right]$$
Using the summation formula of geometric series

\[
G = \frac{1}{f+1} \left\{ \sum_{r=1}^{N} \frac{r\lambda(1-\theta)}{r(1-\theta)\mu + \theta\mu - \lambda} \left( \frac{1}{f} \right)^{k_{f-k_{f-1}}} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f+k_{f-1}}-2k_{0}} \right\}
\]

\[
\times \left[ 1 - \left( \frac{\lambda - r\theta\mu}{r\mu(1-\theta)} \right)^{k_{r-k_{r-1}}} \right] \frac{X_r X_{f-1}}{r\lambda(1-\theta)} \sum_{r=f+1}^{N} \frac{r\lambda(1-\theta)}{r(1-\theta)\mu + \theta\mu - \lambda}
\]

\[
\times \left[ 1 - \left( \frac{\lambda - r\theta\mu}{r\mu(1-\theta)} \right)^{k_{r-k_{r-1}}} \right] \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{r-1}-k_{0}} X_r
\]

\[
+ \sum_{r=1}^{f} \sum_{m=f+1}^{N} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{r-k_{r-1}}+k_{m-k_{m-1}}-2k_{0}} \left[ 1 - \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{m-k_{m-1}}} \right]
\]

\[
\times \left( \frac{(r-m)\lambda^2(1-\theta)^2}{(r(1-\theta)\mu + \theta\mu - \lambda)(m(1-\theta)\mu + \theta\mu - \lambda)} \right) \left[ 1 - \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{m-k_{m-1}}} \right] X_r X_m \}
\]

The first term of \( G \) can be re-written using

\[
\sum_{r=1}^{N} \left[ \frac{r\lambda(1-\theta)}{r(1-\theta)\mu + \theta\mu - \lambda} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}-2k_{0}} \right] \frac{X_r X_{f+1}-X_r X_{f+1}}{r\lambda(1-\theta)} X_{f+1} = -\sum_{r=1}^{N-1} X_r X_{f+1}
\]

\[
\times \left( \frac{\lambda(1-\theta)(\lambda - \mu\theta)}{(r+1)(1-\theta)\mu + \theta\mu - \lambda)(r\mu(1-\theta) + \mu\theta - \lambda) \right)
\]

\[
+ \left( \frac{\lambda(1-\theta)}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}+k_{N-k_{N-1}}-2k_{0}} X_1 X_{f+1} = \frac{N\lambda(1-\theta)}{N(1-\theta)\mu + \theta\mu - \lambda}
\]

\[
\times \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}+k_{N-k_{N-1}}-2k_{0}} X_{N+1} X_{f+1}
\]

The second term of \( G \) can be written as

\[
\sum_{r=f+1}^{N} \left[ \frac{r\lambda(1-\theta)}{r(1-\theta)\mu + \theta\mu - \lambda} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{r-k_{r-1}}-k_{0}} \right] X_{r+1} = \frac{r\lambda(1-\theta)}{r(1-\theta)\mu + \theta\mu - \lambda}
\]

\[
\times \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}+k_{N-k_{N-1}}-2k_{0}} X_{f+1} = \frac{(f+1)\lambda(1-\theta)}{(f+1)\mu - \lambda} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}-k_{0}} X_{f+1}
\]

\[
- \frac{N\lambda(1-\theta)}{N(1-\theta)\mu + \theta\mu - \lambda} \left( \frac{\lambda - \mu\theta}{\mu(1-\theta)} \right)^{k_{f-k_{f-1}}+k_{N-k_{N-1}}-2k_{0}} X_{N+1} = \frac{(1-\theta)(\lambda - \mu\theta)}{(1-\theta)\mu - \lambda}
\]

\[
\times \sum_{r=f+1}^{N-1} \left( \frac{r(1-\theta)\mu + \theta\mu - \lambda((r+1)\mu(1-\theta) + \mu\theta - \lambda) X_{r+1}
\]
The third term of \( G \) can be written as
\[
\sum_{r=1}^{f} \sum_{m=f+1}^{N} \left[ \frac{(r-m)(\lambda(1-\theta))^2 X_r X_m}{(r(1-\theta)\mu + \theta \mu - \lambda)(m(1-\theta)\mu + \mu \theta - \lambda)} \times \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_{r-1}+k_{m-1}-2k_0} - \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_r+k_{m-1}-2k_0} \right]
\]
\[
\times \frac{(r-m)(\lambda(1-\theta))^2 X_{r+1} X_m}{(r(1-\theta)\mu + \theta \mu - \lambda)(m(1-\theta)\mu + \mu \theta - \lambda)}
\]

In a similar way to obtaining the expression of the first two terms, the third term can be written as
\[
G = \frac{1}{f+1} \left[ \frac{\lambda(1-\theta)}{\mu - \lambda} \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_N-k_0} X_{N+1} - \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_f+k_N-2k_0} X_{N+1} \right]
\]
\[
\times \frac{f \lambda(1-\theta)}{f\mu - \lambda} X_{f+1} - \sum_{r=1}^{f-1} \frac{(\lambda(1-\theta))^2}{(r(1-\theta)\mu + \theta \mu - \lambda)((r+1)\mu(1-\theta) + \mu \theta - \lambda)}
\]
\[
\times \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_r+k_N-2k_0} X_{N+1} X_{r+1} \]

There are two cases: (i) \( f = 1 \) (notice that \( X_1 = X_2 = 1 \))
\[
G = \frac{1}{f+1} \left[ \frac{\lambda(1-\theta)}{\mu - \lambda} \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_N-k_0} X_{N+1} - \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_f+k_N-2k_0} X_{N+1} \right]
\]
\[
\times \frac{\lambda(1-\theta)}{\mu - \lambda} X_{N+1} \]

(ii) \( f > 1 \) and \( y = (r(1-\theta)\mu + \theta \mu - \lambda), z = ((r+1)\mu(1-\theta) + \mu \theta - \lambda) \)

Since \( \frac{1}{yz} = \frac{1}{\mu} \left[ \frac{1}{y} - \frac{1}{z} \right] \).
\[
G = \frac{1}{f+1} \left[ \frac{\lambda(1-\theta)}{\mu - \lambda} \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_N-k_0} \left( 1 - \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_f-k_0+1} \right) X_{N+1} \right]
\]
\[
+ \frac{\lambda(1-\theta)}{f\mu - \lambda} \left( \frac{\lambda - \mu \theta}{\mu(1-\theta)} \right)^{k_f+k_N-2k_0+1} X_f X_{N+1}
\]
\[
\times \left( 1 - \left( \frac{\lambda}{f\mu(1-\theta)} - \frac{\theta}{(1-\theta)} \right)^{k_f-k_f-1+1} \right) \]
In these two cases, unless $\mu = \lambda$, $G > 0$ and thus $G/E > 0$ and $F(V) \geq F(T)$.

\[(c)\] Since $B = N - F$, $B(V) \leq B(T)$

This Theorem is quite natural. Since in policy $V$ switching from the back room to the front room occurs earlier, the expected number of servers in the front room can increase and the expected waiting time can improve (or remains the same). The following two corollaries are direct results of theorem.

### 3.2 Corollary

Given a number of server $N$, the policy $\hat{T} = \{k_0 = 0, k_1 = 1, \cdots k_{N-1} = R - 1, k_N = R\}$ yields the smallest possible $W_q$ and $B$ and largest possible $F$, $W_q(\hat{T}) = \min_{T \in \mathbb{R}} W_q(T)$, $F(\hat{T}) = \max_{T \in \mathbb{R}} F(T)$, $B(\hat{T}) = \min_{T \in \mathbb{R}} B(T)$. Obviously the queueing system corresponding to $\hat{T}$ is $M/M/N/R$ with feedback.

### 3.3 Corollary

Given a number of server $N$, the policy $\hat{S} = \{k_0 = R - N, k_1 = R - N + 1, \cdots k_{N-1}, k_N = R\}$ yields the smallest possible $F$ and the largest possible $W_q$ and $B$, $W_q(\hat{S}) = \max_{T \in \mathbb{R}} W_q(T)$, $F(\hat{S}) = \min_{T \in \mathbb{R}} F(T)$, $B(\hat{S}) = \max_{T \in \mathbb{R}} B(T)$. The system corresponding to $\hat{T}$ is an $M/M/N/R$ feedback system where customers are served only when more than $R - N$ customers are in the system. When switching policy changes the constraints set also changes and it is difficult to solve the two models $M_1$ and $M_2$. Therefore to solve the models we focus on developing good algorithms.

### 4 Solving Design and Control Problems

Berman et. al [6] proposed a heuristic method for the solution of this problem. This method is based on the theorem and two corollaries. They proposed the notions of eligible type 1 and type 2 components. An eligible type 1 component is a switching point $k_r$ satisfying the condition $|k_r - k_{r-1}| > 1$ and $k_{r-1} \geq 0, r < N$. A switching point $k_r$ is an eligible type 2 component if $|k_{r+1} - k_r| > 1, r < N$. Queueing design models with heuristic approach are discussed in the work of Berman et. al [1]. An improved feasible solution is obtained, if one of the eligible type 1 components of policy $T$ is reduced by
one. An improved infeasible solution is obtained, if one of the eligible type 2 components is increased by one. Eligible type 1 components and eligible type 2 components will further be referred to simply as type 1 and type 2 components, respectively.

4.1 Optimal Control for Model $M_1$

Step 1: Let $T = \hat{S}$
Step 2: If $B(T) < B_e$, stop. The problem is infeasible. Otherwise, let $Imb \cdot W_q = W_q(T)$ and $Imb \cdot T = T$. Set $f = N$.
Step 3: Find the smallest $Y^*$ such that $0 \leq Y^* < f$ and $k_Y^*$ is an eligible type 1 component. If no such $Y^*$ exists, goto step 5. Set $k_Y^* = k_Y^* - 1$. If $B(T) < B_e$, set $f = Y^*$ and goto step 5.
Step 4: If $W_q(T) < Imb \cdot W_q$, let $Imb \cdot W_q = W_q(T)$ and $Imb \cdot T = T$. Go to step 3.
Step 5: Find the smallest $Y^*$ such that $0 \leq Y^* < f$ and $k_Y^*$ is an eligible type 2 component. If no such $Y^*$ exists, goto step 6. Set $k_Y^* = k_Y^* + 1$. If $B(T) < B_e$, repeat step 5. Else goto step 4.
Step 6: Stop and return $Imb \cdot T$ as the best solution.

4.2 Optimal Design and Control for Model $M_2$

Step 1: $N = 1$
Step 2: Consider the two policies $\hat{T}$ and $\hat{S}$. Let $B_{min} = B(\hat{T}), W_{min} = W_q(\hat{T}), B_{max} = B(\hat{S})$ and $W_{max} = W_q(\hat{S})$.
Step 3: If $B_{max} < B_e$ or $W_{min} > W_u$ goto step 5. If $B_{min} \geq B_e$ and $W_{max} \leq W_u$ stop, $N$ is optimal. Let $N_e = N$.
Step 4: Apply $M_1$. For any policy $T^0$, if $B(T^0) \geq B_e$ and $W_q(T^0) \leq W_u$ stop, $N$ is optimal.

5 Computational Results

We present some numerical results using Matlab in order to illustrate the effect of various parameters on the main performance measures. For the chosen values of $\lambda, \mu, B_s, N, R$ and $W_u$, numerical results are given in Tables 1-5. Table 1 clearly shows that, if the probability of customer feedback increases then the expected number of servers in front room increases and the expected waiting time in the queue is also increases. The expected number of servers in the back room decreases. Similarly results are shown in Table 2 and 3. When the arrival rate and service rate increases, the expected waiting time decreases is shown in Table 4. For increasing values of feedback probability, the mean
Feedback control model

\[ \begin{array}{cccccccccc}
\theta & 0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\
F & 2.663 & 2.733 & 2.797 & 2.854 & 2.902 & 2.940 & 2.967 & 2.986 & 2.996 & 2.999 \\
B & 0.337 & 0.267 & 0.203 & 0.146 & 0.098 & 0.060 & 0.032 & 0.014 & 0.004 & 0.0005 \\
W_q & 0.306 & 0.412 & 0.562 & 0.783 & 1.272 & 1.707 & 2.793 & 5.187 & 10.177 & 15.876 \\
\end{array} \]

Table 1: \( F, B \) and \( W_q \) versus \( \theta \) for model \( M_1 \)

\[ \begin{array}{cccccccccc}
\theta & 0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \\
\lambda = 10-M_1 & 0.202 & 0.192 & 0.179 & 0.165 & 0.149 & 0.131 & 0.110 & 0.088 & 0.068 & 0.049 \\
\lambda = 20-M_1 & 0.185 & 0.198 & 0.213 & 0.232 & 0.256 & 0.289 & 0.334 & 0.411 & 0.509 & 0.714 \\
\lambda = 50-M_1 & 0.186 & 0.215 & 0.252 & 0.300 & 0.363 & 0.451 & 0.582 & 0.799 & 1.226 & 2.459 \\
\lambda = 10-M_2 & 1.9 & 1.792 & 1.690 & 1.594 & 1.503 & 1.418 & 1.34 & 1.267 & 1.202 & 1.146 \\
\end{array} \]

Table 2: \( W_q \) versus \( \theta \) for varying arrival rate

waiting time in the queue increases and the expected number of servers in the back room decreases is shown in Table 5.

Two dimensional graphs are plotted for both the models. For increasing values of service rate, the expected number of servers in the front room decreases is shown in Figure 1. The mean waiting time in the queue decreases for increasing the values of service rate is plotted in Figure 2. For different values of arrival rate, the mean waiting time in the queue increases for increasing values of feedback probability \( \theta \) for model \( M_1 \) is shown in Figures 3 and 7. When the arrival rate increases, the expected waiting time increases is shown in Figures 5-6 for model \( M_1 \) and \( M_2 \) respectively. For varies values of service rate, the mean waiting time in the queue decreases for increasing values of feedback probability \( \theta \) for model \( M_2 \) is shown in Figures 4 and 8.

When the arrival rate increases, the expected waiting time increases is shown in Figure 9 for model \( M_1 \) and \( M_2 \). For model \( M_1 \) and \( M_2 \), the expected waiting time is decreases for varying values of service rate is plotted in Figure 10. Increasing the arrival and service rate, the mean waiting time for queue design and queue control model \( M_2 \) is reduced more when compared with queue control model \( M_1 \). The expected waiting time in the queue for increasing values
of feedback probability $\theta$ and service rate $\mu$ for model $M_1$ and $M_2$ is shown in surfaces Figures 11-12 displays a upward trend as expected. The conclusion of the computational experiment shows that model $M_2$ works extremely well.

### 6 Conclusion

We analyzed the optimal service policy, which controls the number of servers in the front room depending on the number of customer waiting for service. After completing the front room service, if the customer is unsatisfied with the service, he may rejoin the service again until the service is completed successfully or leave the service area forever which was not discussed in the literature so far. Specific switching policy optimizes the expected waiting time in the queue. We analyzed both queue design and control models, which is useful for industrial and manufacturing systems. A brief numerical analysis is done for the performance measures.
Feedback control model

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_q-M_1$</td>
<td>0.290</td>
<td>0.380</td>
<td>0.500</td>
<td>0.670</td>
<td>0.925</td>
<td>1.338</td>
<td>2.081</td>
<td>3.656</td>
<td>8.072</td>
<td>29.474</td>
</tr>
<tr>
<td>$W_q-M_2$</td>
<td>1.143</td>
<td>1.174</td>
<td>1.215</td>
<td>1.269</td>
<td>1.342</td>
<td>1.443</td>
<td>1.589</td>
<td>1.815</td>
<td>2.202</td>
<td>3.000</td>
</tr>
<tr>
<td>$B-M_1$</td>
<td>0.298</td>
<td>0.251</td>
<td>0.207</td>
<td>0.168</td>
<td>0.131</td>
<td>0.098</td>
<td>0.067</td>
<td>0.041</td>
<td>0.019</td>
<td>0.005</td>
</tr>
<tr>
<td>$B-M_2$</td>
<td>0.438</td>
<td>0.429</td>
<td>0.418</td>
<td>0.404</td>
<td>0.387</td>
<td>0.365</td>
<td>0.338</td>
<td>0.303</td>
<td>0.257</td>
<td>0.196</td>
</tr>
</tbody>
</table>

Table 5: $W_q$, $B$ versus feedback probability $\theta$

**ACKNOWLEDGEMENTS.**

This research work has been supported by UGC Research Fellowship - SRF, Government of India.

**References**


**Received: August, 2011**

![Figure 1: F vs service rate](image1)

![Figure 2: Wq vs service rate](image2)
Figure 3: \( W_q \) vs feedback for \( M_1 \)  

Figure 4: \( W_q \) vs feedback for \( M_2 \)  

Figure 5: \( W_q \) vs arrival rate for \( M_1 \)  

Figure 6: \( W_q \) vs arrival rate for \( M_2 \)  

Figure 7: \( W_q \) vs feedback for \( M_1 \)  

Figure 8: \( W_q \) vs feedback for \( M_2 \)
Figure 9: $W_q$ vs arrival rate

Figure 10: $W_q$ vs service rate

Figure 11: $W_q$ vs $\theta$ and $\mu$ for $M_1$

Figure 12: $W_q$ vs $\theta$ and $\mu$ for $M_2$