**RJ10420 (A0710-034) October 24, 2007**
**Other**

# IBM Research Report

## Optimizing the Point-of-Sale Device Mix in a Retail Store

**Ying Tat Leung[1], Mayank Sharma[2], Anupam Singh[3], Myron Flickner[1]**

[1]IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

[2]IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

[3]IBM Retail EBO
12902 Federal Systems Park Drive
Fairfax, VA 22033

**Research Division**
**Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich**

# Optimizing the Point-of-Sale Device Mix in a Retail Store

Ying Tat Leung[1], Mayank Sharma[2], Anupam Singh[3], Myron Flickner[1]

[1]IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

[2]IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

[3]IBM Retail EBO
12902 Federal Systems Park Dr,
Fairfax, VA 22033

October 2007

## Abstract

In recent years, several different types of retail point-of-sale (POS) devices have been developed, in addition to the traditional, fixed registers. They include self-checkout devices and mobile devices which can be installed quickly for temporary use during peak retail seasons. A retail store has to determine the number and type of POS devices it needs in order to serve its customers satisfactorily. We develop an optimization model to find the mix of devices that results in the lowest total cost of owning and operating the devices over their life span. The model was applied to help an IBM client plan a POS device renewal in its chain of 400 stores.

## 1. Introduction

Customer checkout registers, known generally as point-of-sale (POS) devices, represent an important area for a retailer. In many retailers, such as discount stores or grocery supermarkets, the POS area is the only customer touch point inside the store. The experience of a customer there may have a significant influence on the overall impression on the store and even the entire store chain. At the same time, the POS area is a costly investment in the store, in terms of capital (hardware and software) and variable expenses (labor and maintenance costs).

POS devices have become quite sophisticated over the years, growing from independent mechanical cash registers to networked, PC-based devices with all sorts of attachments. In the recent decade, innovations have led to different types of POS devices beyond the common, fixed registers, most notably self-checkout devices, mobiles devices, and portable devices known as "line busters". Self-checkout devices save store labor since customers scan items themselves and therefore one store associate can monitor multiple checkout devices. Mobile devices are designed for easy setup so that they can be added to a store for busy time periods such as holiday seasons and put away otherwise to maximize selling floor space. Line busters are portable, wireless devices that are used to pre-scan items while customers are waiting in line to increase throughput of a POS lane. As can be seen, the motivation for all these innovations is to reduce cost and/or to maintain or improve customer service.

One of the most critical aspects of customer service at a POS device is the customer waiting time in the queue (see, e.g., Houston et al. 1998 and banking customer survey results in Brickstream Corp. 2004). Increasing the number of active POS devices at any time obviously reduces the waiting time, but increases the labor cost and possibly the capital cost of equipment and the opportunity cost of selling space. Because customers arrive randomly at a rate varying over time and their basket sizes are also uncertain, planning for the number of each type of POS devices is not a trivial task. The key decisions are the number of each type of POS devices to install in a store and the number of devices in each type that are open (i.e., manned and serving customers) at any time. The two decisions are clearly related but are usually considered separately because they are encountered in different contexts in practice. The number and type of POS devices to install are infrequent, capital equipment decisions that need to be made once every few years or when a new store is opened. These are what we will call design decisions in this paper. Historically the time interval of this decision was even longer (e.g., 10 years or more), but recent technological innovations (as discussed above) have shortened the time interval. Decisions on what devices to open over the course of a day, given a fixed set of devices installed at the store, are largely made as staff scheduling is done, usually daily. To a more limited extent, real time decisions can be made to adjust the number of open devices scheduled as store traffic varies from that expected. We will denote such daily or real-time decisions by operation decisions.

The queueing system at a typical bank of POS devices is, strictly speaking, rather complicated. It is generally known that the customer inter-arrival time distribution is time inhomogeneous. The service time distribution may also be time inhomogeneous due to the average basket sizes varying over a day. Price checks and other issues may add significant time to some customers, almost acting like a server breakdown. Customers may abandon their basket and leave the queue or choose not to join the queue at all (and leave the store or continue shopping). In the case of individual queues in front of each POS device, customers may jockey between queues depending on which queue is perceived to be moving quickly. An exact queueing analysis is therefore quite difficult. To our knowledge such an analysis does not exist in the research literature. In addition, most retailers do not have data to fully characterize the shopping basket arrival times at the POS queue. Unless one uses very sophisticated video imaging technology, the arrival times at the POS queue have to be collected manually over a significant period of time (to account for time inhomogeneity) at the stores. Such studies are relatively expensive and are at best done very infrequently.

Even though the POS queue appears to be a technically interesting and practically useful subject to study, few papers have been published on it. Metz and Savir (1975) developed a simulation model and an M/G/c queueing model to analyze the performance of POS devices in supermarkets. The M/G/c model was used to estimate number of devices needed to achieve a given throughput that is below the point of overload, and the simulation model is used to estimate the performance at and above the point of overload. These models are useful for POS design decisions. Williams et al. (2002) developed a simulation model to help determine the critical queue length beyond which a new POS device should be open. This analysis helps set a policy for the real-time operation decision.

## 2. A General Device Mix Optimization Problem

In this study, we are interested in minimizing the total cost of owning and operating the POS devices in a selected retail store, given a performance target for the POS queue and possibly other business constraints. We consider the possibility of installing and using three types of POS devices: fixed checkout registers, self checkout registers, and mobile checkout devices. Line busters are not explicitly considered but they can be added to the formulation of the general problem easily, just like a mobile checkout device. In fact, line busters and mobile checkout devices are similar in the sense that they can be added to the permanently installed devices (fixed and self-checkout devices) in a store temporarily. In the next section, we will treat mobile and line busters similarly in solving a practical version of the optimization problem.

The following notation will be used.

**Decision variables:**

$N_1$ = number of self checkout (SCO) registers to be purchased

$N_2$ = number of fixed checkout registers to be purchased

$N_3$ = number of mobile checkout devices to be purchased

$N_{1t}$ = number of self checkout registers available in time period $t$ (i.e., number of self checkout registers that are open in period $t$)

$N_{2t}$ = number of fixed checkout registers available in time period $t$ (i.e., number of fixed checkout registers that are open in period $t$)

$N_{3t}$ = number of mobile checkout devices available in time period $t$ (i.e., number of mobile checkout devices that are open in period $t$)

**Parameters:**

$F_1$ = Per time period cost of owning one self checkout register, including amortized purchase cost over its useful life, maintenance costs, real estate cost to accommodate the register (fixed costs)
$F_2$ = Per time period cost of owning one fixed checkout register
$F_3$ = Per time period cost of owning one mobile checkout device
$V_1$ = Per time period cost of operating one self checkout register, including energy costs, labor cost to operate the register (variable costs)
$V_2$ = Per time period cost of operating one fixed checkout register
$V_3$ = Per time period cost of owning and operating one mobile checkout device
$\lambda_t$ = total arrival rate of shopping baskets to all the checkout registers in time period $t$
$\mu_1$ = service rate of one self checkout register = 1/(mean service time per transaction at a self checkout register)
$\mu_2$ = service rate of one fixed checkout register = 1/(mean service time per transaction at a fixed checkout register)
$\mu_3$ = service rate of one mobile checkout device = 1/(mean service time per transaction at a mobile checkout device)
$\pi_t$ = service level requirement in time period t in terms of fraction of customers having a waiting time of $W_{0t}$ or smaller
$W_{0t}$ = waiting time threshold in time period $t$
$Q_{0t}$ = queue length threshold in time period $t$
$m_1$ = lower bound of number of self checkout registers open at any time
$m_2$ = lower bound of number of fixed checkout registers open at any time
$M_1$ = upper bound of number of self checkout registers open at any time
$M_2$ = upper bound of number of fixed checkout registers open at any time

**Queueing sub-model:**
$g(\lambda_t, N_{1t}, N_{2t}, N_{3t})$ = expected number of customers arriving in time period $t$ with waiting times $\leq W_{0t}$
$h(\lambda_t, N_{1t}, N_{2t}, N_{3t})$ = expected number of customers waiting in queue in time period $t$

**Internal variables:**
$t$ = time period in which the arrival rate of shopping baskets is constant =1, 2, …, $T$
Without loss of generality, assume the length of each time period = 1 (We define the time scale such that each time period is 1 time unit)
$n_t$ = 0, 1, 2, …

To minimize the total cost of owning the set of POS devices for a time horizon of $T$ periods, we have the following general formulation.

**(M1) Optimization model:**

(1.1) $$\min C = F_1TN_1 + F_2TN_2 + F_3TN_3 + V_1\sum_{t=1}^{T} N_{1t} + V_2\sum_{t=1}^{T} N_{2t} + V_3\sum_{t=1}^{T} N_{3t}$$

S.T.
(1.2) $N_{1t} \leq N_1$ , $t$=1, 2, ..., $T$
(1.3) $N_{2t} \leq N_2$ , $t$=1, 2, ..., $T$

(1.4) $N_{3t} \le N_3$ , $t$=1, 2, ..., $T$

(1.5) $\sum_{t=1}^{T} \pi_t \lambda_t \le \sum_{t=1}^{T} g(\lambda_t, N_{1t}, N_{2t}, N_{3t})$

(1.6) $h(\lambda_t, N_{1t}, N_{2t}, N_{3t}) \le Q_{0t}$ , $t$=1, 2, …, $T$

(1.7) $m_1 \le N_{1t} \le M_1$ , $t$=1, 2, ..., $T$

(1.8) $m_2 \le N_{2t} \le M_2$ , $t$=1, 2, ..., $T$

(1.9) $0 \le N_{3t}$ , $t$=1, 2, ..., $T$

(1.10) $N_{1t} = 4n_t$, $t$=1, 2, ..., $T$

(1.11) $n_t, N_{2t}, N_{3t}$ = 0, 1, 2, …

Objective function (1.1) represents the total cost of ownership of all the checkout devices over a planning horizon $T$. $T$, for example, can be one year and (1) will be the annual cost of ownership.
Constraints (1.2)-(1.4) set $N_i = \max_t N_{it}$, so that the fixed cost is accounted for the number of registers installed.

Constraints (1.5)-(1.6) represent the required service level, expressed in the expected fraction of customers with waiting times ≤ $W_0$ and the expected number of customers waiting in line, respectively. Note that in constraint (1.5), in general the expected fraction of customers may not be equivalent to the probability of waiting time ≤ $W_0$. (For ergodic systems they are equivalent.) Most likely only one of the two constraints will be active in the final solution.

Constraints (1.7)-(1.8) represent business rules, such as at least one fixed register is always open, and a limit of the number of registers that can be installed in a store, imposed by physical and business constraints (e.g., maximum space allocated to checkout).

Constraint (1.9) assumes that there is no upper limit on the number of mobile devices since they do not occupy any physical space permanently.

Constraint (1.10) represents the fact that self checkouts typically come in pods of four at the time of this study.

$g(\lambda_t, N_{1t}, N_{2t}, N_{3t})$ and $h(\lambda_t, N_{1t}, N_{2t}, N_{3t})$ depend on the specific queueing model of the POS devices. Note that even for the simplest queueing models, $g$ and $h$ will be nonlinear functions. Other customer service criteria, based on common performance measures of a queue, are also possible. Typically we assume that in each time period the parameters of the queueing sub-model (e.g., the arrival rate) are stationary. This has been studied as the pointwise stationary approximation of a time-varying queueing system (Green and Kolesar 1991).

# 3. A Practical Version of the General Problem and Its Solution

Based on our knowledge of how retailers typically use POS devices and the specific requirements of one retailer, we develop the following version of the optimization problem.

Continuing to use the same notations as those in the prior section, we further let

$\lambda_{12t}$ = arrival rate of shopping baskets to the SCO and fixed POS registers in time period $t$
$\lambda_{3t}$ = arrival rate of shopping baskets to the mobile POS registers in time period $t$

$\psi$ = maximum allowable ratio of SCO registers to the sum of SCO and regular POS registers

## 3.1. The Queueing Model

We model the bank of POS devices (including SCO, regular, and mobile POS devices) as a multi-server queue with Poisson arrivals and exponential service times (M/M/$k$ queue). Based on observations in a retail store, Williams et al. (2002) found that inter-arrival times at the POS queue can adequately be characterized by a Poisson distribution with rates dependent on the time of day and day of week. Metz and Savir (1975) also used a Poisson distribution for the arrival process. The service process consists of three different parts: an initiation (greeting the customer, scanning a customer loyalty card, etc.), scanning the items, and payment tendering. The initiation time is usually rather small, since the POS device will automatically initiate a new transaction by scanning the first item. The time for item scanning obviously depend on the number of items in the shopping basket and payment tendering time depends on the type of payment used (cash, credit card, check, etc.) Williams et al. (2002) found from their empirical data that the exponential distribution is a fairly good approximation for modeling the overall service time.

For the entire bank of POS devices, we use a multi-server queue model where a single queue is formed in front of multiple POS devices. This could be an accurate description for some cases, e.g., most banks and airline check-in counters. In other cases, such as a grocery store, customers form individual lines in front of every open POS device. A customer will jockey to another line if that line becomes shorter than his current position. In particular, when a device becomes available (i.e., no one is waiting), customers from other lines will move into that device, such that whenever there are some customers waiting across all devices, no device will be empty. The multi-server model satisfies this boundary condition. When all servers are busy, the multi-server model is an approximation to the actual individual queues with jockeying. Metz and Savir (1975) made the same observations but also used the multi-server queue in their analytic model.

Williams et al. (2002) captured individual queues with jockeying in their simulation model.

We take the weighted average service rate, i.e., $(N_1\mu_1 + N_2\mu_2 + N_3\mu_3) / (N_1 + N_2 + N_3)$, as the service rate of any server in the identical bank of servers in our standard multi-server queue model. This is an approximation to reality where the service rates of the different types of POS devices are not the same. In practice, the service rates of fixed and mobile POS devices are very close. Self-checkout devices may be quite different, representing a source of potential accuracy improvement.

## 3.2. Calculation of Mobile POS Devices

Because mobile POS devices cannot handle all tender types (e.g., they cannot handle cash), they are only used to supplement SCO and regular POS devices during periods of peak traffic. Hence, the number of mobile POS devices is calculated by assigning to them a user-defined peak portion of the POS traffic and using the waiting time criterion.

Let
$\alpha$ = percentile traffic parameter such that the least congested $(1 - \alpha)$ fraction of the time periods are to be handled without mobile devices
$\lambda_{(1-\alpha)}$ = the $(1 - \alpha)$-th percentile arrival rate in $\{\lambda_t, t = 1, 2, \ldots, T\}$
$\theta_t$ = fraction of arrivals with tender types that a mobile device can take, in time period $t$

$\Lambda^+ = \{t: \lambda_t \geq \lambda_{(1-\alpha)}\}$
$\Lambda^- = \{t: \lambda_t < \lambda_{(1-\alpha)}\}$

Then

(2.1) $$\lambda_{3t} = \begin{cases} \min(\lambda_t - \lambda_{(1\text{-}\alpha)}, \lambda_t\theta_t) & ,t \in \Lambda^+ \\ 0 & ,t \in \Lambda^- \end{cases}$$

Dropping the subscript $t$, let $\lambda$ denote the arrival rate to the bank of POS devices (modeled as an M/M/$k$ queue) and $\mu$ denote the service rate of one server in this queue. It is known (e.g., Gross and Harris (1998), Section 2.3) that

(2.2) $$\text{P}\{\text{waiting time of a random customer} \leq W\} = 1 - p_{02}\, p_0 e^{-(k\mu-\lambda)W},$$

where

$$r = \frac{\lambda}{\mu}$$

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

$$p_{01} = \sum_{n=0}^{k-1} \frac{r^n}{n!}$$

$$p_{02} = \frac{r^k}{k!(1-\rho)}$$

$$p_0 = \frac{1}{p_{01} + p_{02}}$$

For $t$=1, 2, …, $T$, we solve for the smallest $N_{3t}$ by setting $k = N_1^* + N_{2t}^* + N_{3t}$ and $\mu = (N_1^*\mu_1 + N_{2t}^*\mu_2 + N_{3t}\mu_3) / (N_1^* + N_{2t}^* + N_{3t})$ in the probability calculation (2.2), i.e., we solve for the smallest $N_{3t}$ such that

(2.3) $\pi_t \leq 1 - p_{02} p_0 e^{-(k\mu-\lambda)W}$ ,

and taking $N_1^*$ and $N_{2t}^*$ to be the solution obtained in solving the problem described in Section 3.3. Because (2.2) is monotonic in $k$ and in practice the number of mobile devices is usually small (e.g., <10), a simple linear search, starting from 1, is used to solve for the smallest $N_{3t}$ for each $t$.

Finally,

(2.4) $N_3^* = \max_t N_{3t}$, is the solution for the number of mobile devices.

## 3.3. Calculation of Self Checkout and Fixed POS Devices

An optimization model is developed to find the number of SCO and fixed POS devices such that the total cost per year is minimized and the service level requirements are satisfied. This model is a variation of the general problem stated in Section 2.

The following assumptions are used in formulating the optimization model.

1. Service level requirement is applied per time period, i.e., the service level requirement is met in each individual time period.
2. In each time period, the smallest number of devices within the bound of available devices satisfying the service level requirement are open. This means that the operation decisions of when to open a POS device are optimal.
3. Self checkout registers are always open.

The optimization model is then as follows.

(M2) Optimization Model

(3.1) $\min C = F_1 T N_1 + F_2 T N_2 + V_1 T N_1 + V_2 \sum_{t=1}^{T} N_{2t}$

S.T.

(3.2) $N_{2t} \leq N_2$ , $t = 1, 2, ..., T$
(3.3) $\pi_t \leq$ P{waiting time of a random customer in period $t \leq W_{0t}$}, $t = 1, 2, \ldots, T$
(3.4) $m_1 \leq N_1 \leq M_1$
(3.5) $m_2 \leq N_{2t} \leq M_2$ , $t = 1, 2, ..., T$
(3.6) $N_1/(N_1 + N_2) \leq \psi$
(3.7) $N_1 = 4n$
(3.8) $n, N_{2t} = 0, 1, 2, \ldots, t = 1, 2, ..., T$
(3.9) $\lambda_{12t} = \lambda_t - \lambda_{3t}$, $t = 1, 2, ..., T$

Objective function (3.1) represents the total cost of ownership of all the checkout devices over a planning horizon $T$. As mentioned, we assume that SCO devices (if any) are open in every time period.

Constraint (3.2) set $N_2 = \max_t N_{2t}$, so that the fixed cost is accounted for the number of devices installed.

Constraints (3.3) represents the required service level, expressed in probability of a random customer with waiting time $\leq W_{0t}$. This probability is dependent on the specific queueing model used to represent the POS devices. The constraint enforces the service level requirement on an epoch-by-epoch basis, i.e., the service level requirement is enforced independently for every time period.

To calculate the RHS of (3.3), all devices together (SCO and fixed) are modeled as a single M/M/$k$ queue (as discussed above) with $k_t$ identical servers, where $k_t = N_1 + N_{2t}$. The service rate of each server, $\mu_t$, is calculated by $\mu_t = (N_1\,\mu_1 + N_{2t}\,\mu_2) / (N_1 + N_{2t})$. Then, P{waiting time of a random customer in period $t \leq W_{0t}$}

$$= 1 - \frac{r_t^{k_t} p_0}{k_t!(1-\rho_t)} e^{-(k_t\mu_t - \lambda_t)W_{0t}} ,$$

where $r_t = \lambda_t / \mu_t$ , $\rho_t = \lambda_t / (k_t\,\mu_t)$ ,

$$p_0^{-1} = \left( \sum_{n=0}^{k_t-1} \frac{r_t^{\,n}}{n!} \right) + \frac{r_t^{k_t}}{k_t!(1-\rho_t)}$$

Hence, (3.3) becomes

(3.3') $$\pi_t \leq 1 - \frac{r_t^{k_t} p_0}{k_t!(1-\rho_t)} e^{-(k_t\mu_t - \lambda_t)W_{0t}}$$

Constraints (3.4)-(3.5) represent business rules, such as at least one fixed device is always open, and a limit on the number of devices that can be installed in a store, imposed by physical and business constraints (e.g., maximum space allocated to checkout).

Constraint (3.6) represents another business rule, stating that the number of SCO devices may be limited to a certain fraction, to account for the fact that some consumers are not willing to use SCO devices.

Constraint (3.7) represents the fact that SCO devices come in pods of four.

Constraint (3.9) calculates the arrival rate to the SCO and regular POS devices without the mobile POS devices, where $\lambda_{3t}$ is calculated in (2.1) above.

Noting that the arrival rate to the POS devices is finite, $(N_1 + N_{2t})$ is typically small in practice (say ≤ 30), and the expected waiting time is monotonic in the total number of servers $k$, we solve (M2) using a simple linear search algorithm over possible values of $N_1$, as follows.

**Algorithm to solve M2:**

$N_1 = m_1$ (Assume $m_1$ is a multiple of 4)
While ($N_1 \le M_1$) {
    For $t = 1$ to $T$ {
        If $N_1 = 0$ then $N_{2t} = \max(1, m_2)$ else $N_{2t} = \max(0, m_2)$
        $p = 0$
        While ($N_{2t} \le M_2$) and $p < \pi_t$ {
            $k_t = N_1 + N_{2t}$
            $r_t = \lambda_t(N_1 + N_2) / (N_1\mu_1 + N_{2t}\mu_2)$
            $\rho_t = \lambda_t / (N_1\mu_1 + N_{2t}\mu_2)$

$$p_0^{-1} = \left( \sum_{n=0}^{k_t - 1} \frac{r_t^{\,n}}{n!} \right) + \frac{r_t^{\,k_t}}{k_t!(1-\rho_t)}$$

$$p = 1 - \frac{r_t^{\,k} p_0}{k!(1-\rho_t)} e^{-(k\mu_t - \lambda_t)W_{0t}}$$

            $N_{2t} = N_{2t} + 1$
        }
        $N_{2t} = N_{2t} - 1$
    }
    $K = \max_t \{k_t\}$
    $N_2 = \max_t\{ N_{2t}\}$
    Apply any other user specified constraints to $N_2$ (e.g. ratio of $N_1$:$N_2$)
    If $(N_1 + N_2) \ge K$ {

$$C_{N_1} = F_1TN_1 + F_2TN_2 + V_1TN_1 + V_2\sum_{t=1}^{T} N_{2t}$$

    } else {

$$C_{N_1} = \infty$$

    }
    $N_1 = N_1 + 4$
}
$C^* = \min_i\{C_i\}$ and the corresponding $N_1^*$, $N_2^*$, $N_{2t}^*$ ($t$=1, 2, …, $T$) are the optimal solution

[End of Algorithm]

### 3.4. Overall Solution

The overall solution is $N_3 = N_3^*$ obtained from Section 3.2, and $N_1 = N_1^*$, $N_2 = N_2^*$, $N_{2t} = N_{2t}^*$ obtained from Section 3.3. The total cost is

$$C = F_1TN_1 + F_2TN_2 + F_3TN_3 + V_1TN_1 + V_2\sum_{t=1}^{T} N_{2t} + V_3\sum_{t=1}^{T} N_{3t} \quad (4.1)$$

## 4. Some Further Consideration of Self-Checkout Devices

One of the simplifying assumptions we made in Section 3 is that self-checkout POS devices are always open. This was made primarily for the purpose of easy computation since we now do not have to search over all possible combinations of open fixed and self-checkout devices. It did not pose a problem in practice because many stores would turn out to use one pod of self-checkout and it is indeed least costly to open the self-checkout first, given that it is already installed and that there is no other over-riding business constraints.

In cases when the solution calls for more than one self-checkout pods, it may not be necessary to open all self-checkout pods in every time period and the algorithm will not provide the exact least cost solution. Using again the observation that it is least costly to open self-checkout devices than regular devices (at one quarter the labor rate and assuming that the maintenance costs are comparable), given a certain number of self-checkout and fixed devices, it is optimal to open the least number of self-checkout devices and open the least number of additional fixed devices after all self-checkouts are opened. This logic can be added to a two dimensional version of the linear search algorithm given in Section 3.3 and can save significant amount of effort when compared to an exhaustive search of all combinations of open fixed and self-checkout devices.

## 5. A POS Device Optimization Tool

A POS Device Optimization Tool was developed to solve the optimization problem discussed in Section 3. The tool was written as a standalone Microsoft Excel application using a combination of spreadsheet formulae and Visual Basic in Excel. Figure 5.1 shows the main user screen in the tool. Details of the tool are described in Leung et al. (2007).

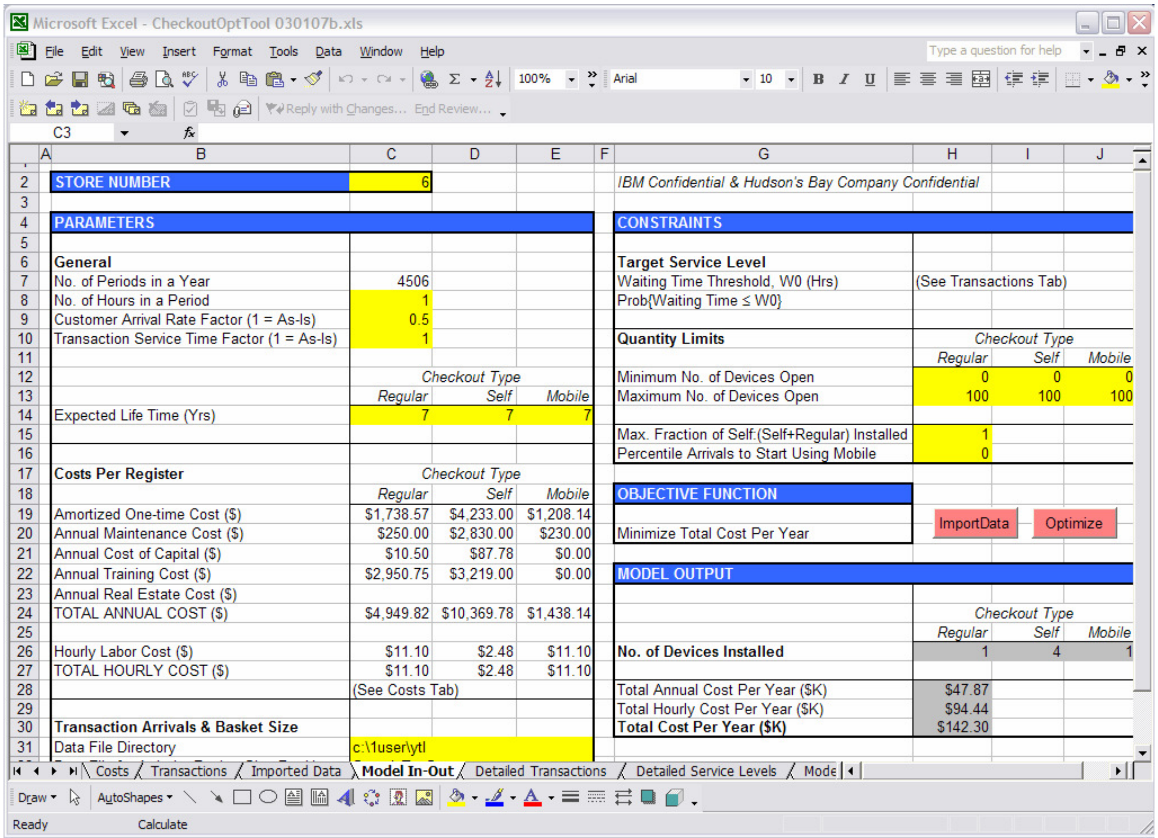


Figure 5.1. Main User Screen of POS Device Optimization Tool

The tool was designed for interactive use such that what-if analysis can be made in real time to facilitate POS design decisions. For example, by varying the waiting time threshold, one can gain insight into how the total cost per year will increase as the waiting time threshold is reduced. With knowledge on how customers perceive waiting times at checkout, a good, if not optimal, operating point in the waiting time – cost space can be chosen.

Since each store has its own traffic pattern, the POS device optimization problem has to be solved for each store independently. The tool provides Visual Basic functions that can be called by a higher level VB routine to perform batch processing. It has been used in such a way by an IBM client to solve the device optimization problem for 400 stores.

# 6. A Numerical Example

We present a real-life example where we used the POS Device Optimization Tool to find the number of devices for a retail chain of general merchandise discount stores. The example is on a selected store which has two separate banks of checkout registers in two different places (i.e., two different entrance/exit areas). The business problem is to determine how many checkout devices of each type (self-checkout, fixed, and mobile) should be installed in each bank in order to minimize the total cost of owning and operating the POS devices per year over the expected life span (taken to be 7 years).

The total cost includes the cost of purchase and install, yearly maintenance cost (under a fixed cost maintenance contract), staff training cost, and hourly labor cost of operating the POS devices. The last item depends on when the device is open, while the rest depend on how many devices are installed. In this particular example, the mobile devices are assumed to be able to take only one form of payment which is credit cards.

Historical checkout data were collected and analyzed to provide the input parameters to the queueing sub-model. Based on the data pattern, we chose to assume that arrival rates to the queue are constant within an hour. The arrival rate to each bank of POS devices is approximated by half of the historical checkout completion rate of the entire store. The mean service time is taken to be
(6.1) Mean service time = (Average number of items in basket)(Mean time to checkout an item) + (Mean time to tender payment).

The average number of items in a basket was estimated from historical data and the mean time to checkout an item and was taken from time standards established by the retailer. The mean time to tender payment depends on the type of payment the customer uses (e.g., cash, credit card, debit card, check, etc.) The retailer has established standard times to handle each payment type and the overall mean is calculated using the fraction of each payment type collected from historical data. Note that all three quantities on the right hand side of equation (6.1) can vary by the hour, similar to the arrival rate.

Figures 6.1 – 6.3 show the results of the optimization model discussed in Section 3, with three different peak portions of POS traffic assigned to mobile devices. For example, Figure 6.1 shows the optimization results as the customer service criteria (threshold for customer waiting time and probability of a customer’s waiting time below the threshold) vary. The red line shows the lowest cost solution and the green line shows the highest cost solution. These results clearly show the incremental cost of improvements in the customer service criteria. They enable a retail executive to find an optimal tradeoff between cost and customer service for their particular business.

| Store Number | Number of Banks | Percentile to start using Mobile | Number of SCOs | Number of Regular POS | Number of Mobile POS | Wait Time (W0) in minutes | Probability Customer Waits less than W0 | Annual Cost ($K) | Annual Sales ($K) |
|---|---|---|---|---|---|---|---|---|---|
| 453 | 2 | 1 | 4 | 7 | 0 | 1 | 0.85 | $361.78 | $26,477 |
| 453 | 2 | 1 | 4 | 7 | 0 | 3 | 0.85 | $313.04 | $26,477 |
| 453 | 2 | 1 | 4 | 6 | 0 | 5 | 0.85 | $289.92 | $26,477 |
| 453 | 2 | 1 | 8 | 7 | 0 | 1 | 0.99 | $459.36 | $26,477 |
| 453 | 2 | 1 | 4 | 8 | 0 | 3 | 0.99 | $396.92 | $26,477 |
| 453 | 2 | 1 | 4 | 7 | 0 | 5 | 0.99 | $346.70 | $26,477 |

Figure 6.1. Optimization Results for the Case of No Mobile Devices

| Store Number | Number of Banks | Percentile to start using Mobile | Number of SCOs | Number of Regular POS | Number of Mobile POS | Wait Time (W0) in minutes | Probability Customer Waits less than W0 | Annual Cost ($K) | Annual Sales ($K) |
|---|---|---|---|---|---|---|---|---|---|
| 453 | 2 | 0.9 | 4 | 6 | 2 | 1 | 0.85 | $363.32 | $26,477 |
| 453 | 2 | 0.9 | 4 | 5 | 2 | 3 | 0.85 | $303.80 | $26,477 |
| 453 | 2 | 0.9 | 4 | 5 | 2 | 5 | 0.85 | $291.66 | $26,477 |
| 453 | 2 | 0.9 | 8 | 5 | 2 | 1 | 0.99 | $448.32 | $26,477 |
| 453 | 2 | 0.9 | 4 | 6 | 2 | 3 | 0.99 | $387.52 | $26,477 |
| 453 | 2 | 0.9 | 4 | 6 | 2 | 5 | 0.99 | $347.06 | $26,477 |

Figure 6.2. Optimization Results for the Case of 10% Checkout Traffic Allocated to Mobile Devices

| Store Number | Number of Banks | Percentile to start using Mobile | Number of SCOs | Number of Regular POS | Number of Mobile POS | Wait Time (W0) in minutes | Probability Customer Waits less than W0 | Annual Cost ($K) | Annual Sales ($K) |
|---|---|---|---|---|---|---|---|---|---|
| 453 | 2 | 0.8 | 4 | 6 | 2 | 1 | 0.85 | $367.28 | $26,477 |
| 453 | 2 | 0.8 | 4 | 5 | 2 | 3 | 0.85 | $307.60 | $26,477 |
| 453 | 2 | 0.8 | 4 | 5 | 2 | 5 | 0.85 | $295.36 | $26,477 |
| 453 | 2 | 0.8 | 8 | 5 | 2 | 1 | 0.99 | $451.24 | $26,477 |
| 453 | 2 | 0.8 | 4 | 6 | 2 | 3 | 0.99 | $391.54 | $26,477 |
| 453 | 2 | 0.8 | 4 | 6 | 2 | 5 | 0.99 | $352.98 | $26,477 |

Figure 6.3. Optimization Results for the Case of 20% Checkout Traffic Allocated to Mobile Devices

## 7. Future Research

As mentioned earlier, research literature on the retail POS queue is rather scarce, leaving a number of opportunities for future research. Two of these are as follows.

1. A more accurate analytical model of the POS queue. Intuitively, the POS queue is somewhere between a standard multi-server queue and a set of single server queues. The multi-server queue satisfies the boundary condition of never having a server idle when there is at least one customer in the queue. A more accurate model going beyond this boundary condition will be useful. This may involve better understanding the behavior of customers, e.g., when they jockey between lines or when they abandon the queue altogether.

2. More comprehensive empirical studies of the arrival and service distributions. Data on service distributions can be obtained from POS transaction logs. Detailed time stamps are often available, even though analyzing the data is not entirely trivial due to the number of possible events in the transaction logs. Data on arrivals at the POS queue are more difficult to obtain. Image-processing based instruments are available but they are not widely used at the present time. The only other alternative is to manually collect data by observing the queues over time. Watching a video of the queue is preferable since one can check the accuracy of the manual data records if the need arises, and the data will be less biased if the video camera can be hidden from view.

## References


Brickstream Corp. (2004). "The value of delivering a better branch experience: Customer market research," white paper.

L. Green and P. Kolesar (1991). "The pointwise stationary approximation for queues with nonstationary arrivals," *Management Science*, Vol. 37, 84-97.

D. Gross and C.M. Harris (1998). *Fundamentals of Queueing Theory*, Third Edition, John Wiley, New York.

M.B. Houston, L.A. Bettencourt, S. Wenger (1998). "The relationship between waiting in a service queue and evaluations of service quality: A field theory perspective," *Psychology & Marketing*, Vol. 15, No. 8, 735-753.

Y.T. Leung, M. Sharma, A. Singh, M. Flickner (2007). "A User's Guide to the Checkout Device Mix Optimization Tool Developed by IBM Corporation for Hudson's Bay Company," project report, March 2007.

W. C. Metz, Jr. and D. Savir (1975). "Store performance studies for the supermarket system," *IBM Systems Journal*, Vol. 14, No. 1, 46-63.

E.J. Williams, M. Karaki, C. Lammers (2002). "Use of simulation to determine cashier staffing policy at a retail checkout," *Proceedings of the 14th European Simulation Conference*, A. Verbraeck, W. Krug (Eds.).