Estimation of a caller retrial rate for a telephone information system

Karla L. HOFFMAN and Carl M. HARRIS
Department of Systems Engineering, George Mason University, Fairfax, VA 22030, U.S.A.

Abstract: As part of a continuing study of the usage of its Taxpayer Service Telephone Network, the U.S. Internal Revenue Service wished to determine more accurate methods for demand measurement. It has long been recognized that the total number of calls coming into such a busy telephone system overestimates the actual number of distinct callers. The Service had previously estimated its real demand by adding \( \frac{3}{4} \) of both the number of blocked or overflow calls and the number of abandonments to the total actually answered. The thrust of this current study then was to develop an accurate statistical method for providing a more objective formula for this true demand, which turns out to be equivalent to estimating the probability of retrial by blocked and abandoned callers.

The major result which has come from this effort is that the average daily retrial percentage taken across location and time of year seems to be moderately stable about a mean value of 69\%, somewhat dependent on both location and (particularly) time of year. The value is consistently higher during periods close to important filing milestones and lower otherwise. We show this to mean that, whenever a rate of 69\% is used, the actual demand would be estimated by augmenting completed loads by 31\% of the number of blocked and abandoned calls for the period of concern.

Keywords: Communications, queues, stochastic processes

1. Introduction

A toll-free telephone system has been established by the U.S. Internal Revenue Service (IRS) to permit taxpayers to obtain tax assistance in a timely and inexpensive manner. The IRS covers the Continental United States by means of a number of strategically located answering sites, which can be reached for a charge no more than that of a local call. The efficient management of these Taxpayer Information Centers is a very difficult problem at best, but it has been hoped that rational approaches can be developed for the necessary resource requirement decisions. The problem of modeling this system was introduced in Harris, Hoffman and Saunders (1984), which portrays the initial simulation study of the IRS network by the Center for Applied Mathematics (CAM) of the National Bureau of Standards (NBS), together with the second or analytic model phase, performed by the Center for Management and Policy Research in conjunction with NBS. (This reference is henceforth called HHS.)

For each site, trunklines must be provided for receiving both local and long-distance calls. The number of lines necessary for any specific location is a function of numerous things, including the rate at which calls come in, the number of servers available, the duration and type of each call, the behavior of customers not served immediately, and so on.

Incoming calls are deemed overflows when all trunklines corresponding to the call type are busy (this results in a busy signal). Any taxpayer who has received a busy signal can opt to dial again (either immediately or some time later) or choose not to retry and instead obtain the information elsewhere. Taxpayers who do not decide to try IRS again are considered as lost.
If, however, all of the answering agents are busy but at least one trunkline is open, then the caller receives the usual kind of prerecorded message. If willing, the taxpayer queues up until his or her turn may come, at which time the call is properly distributed to the available server. It may, of course, be possible that an incoming call may find an idle server, in which case entry is immediate. But if after connection the caller is unwilling to stay on before service begins, an abandon or renege has occurred.

After a caller has received service, there may still remain after-call work for the server. In this event, the line is freed allowing another caller to claim that trunkline, though the server does not receive another caller until the post-hang-up work is finished. Technical details of the elements of the IRS network are given in HHS.

The ultimate issue is for IRS to determine the configuration of staff and equipment at each location which provides the best balance of resources and customer service. The amount of money is essentially fixed by the United States Congress and thus the problem becomes one of ‘satisfying’ demand subject to dollar constraints. However, it has been realized that classical approaches (primarily from the telecommunications literature) to such problems cannot provide adequate guidance to IRS decision makers. As a result, IRS turned to the Center for Applied Mathematics of NBS for some guidelines. For a first step, NBS built a simulation model which allowed for the testing of a variety of alternative approaches to resource management (again see HHS). This is a Monte-Carlo model which was built to replicate the IRS system as closely as possible and thus to permit the investigation of the effects on quality of service of changes in any of the key parameters of the system. Unfortunately, given a veritable infinity of policy choices, it is not feasible for IRS to utilize a computer model for day-to-day decision making. The combinations of alternatives are just too extensive to permit the construction of meaningful tables and too costly to run on an interactive basis whenever needed. The simulation model was built in the first place because it seemed quite clear that an analytic model could not be derived as a very close representation of this network type. Fortunately, detailed data analysis and exhaustive running of the model provided important insights on the effect of model assumptions upon the output. This then opened up the real possibility that closed-form analytic solutions might eventually be possible, and the findings of the NBS/CAM study indeed pointed to a streamlined queue-theoretic version of a typical IRS location.

2. Results of NBS work

The NBS study was quite extensive and led to numerous findings about the IRS system. However, not all are of primary interest to this current effort. We focus then on points made which relate to the ultimate derivation of an approximating analytic model. Judged in this sense, there are six major conclusions of the NBS work (see HHS):

(1) From tradition and the simulation study, the originating input stream can be well modeled by a Poisson process;

(2) From data analysis and simulation, service times (when service and after-hang-up work are combined) are well approximated as a set of independent and identical exponential random variables. Including after-call time does not significantly alter the system’s capability to handle calls;

(3) From observation and simulation, the system builds up very quickly to a statistical equilibrium and thus a stationary queueing analysis should be appropriate;

(4) From simulation, it is adequate to consider calls to be handled on a first-come, first-served basis aggregating local and long-distance calls;

(5) From observation, there is a nontrivial number of customers who abandon or renege after connection but prior to service; and

(6) From observation, there is significant redialing by callers who have previously received busy signals or had abandoned (the estimation of their frequency is the major issue of this paper).

We shall not go through these conclusions in great detail, but rather combine them into what might be called a model of choice. That is to say, the NBS simulation model suggests that the underlying complexity of a Taxpayer Service Center may be well represented (in the sense of matching its measures of effectiveness) by a multi-channel queueing system with basic characteristics as indicated by the six items. Fortunately, as we hoped, this model is indeed solvable, and thus offers an excellent opportunity for a more satisfactory tool
for supporting IRS decision making.

To be more specific on the model choice, it begins from a Poisson input stream feeding one queue and serviced by numerous independent and identical exponential servers. There is a limited waiting area (say of size $K$) and any customer finding all waiting spaces filled cannot enter and must either redial or opt not to return. Customers enter service without priority in a first-come, first-served manner (FCFS), and may decide to abandon before service if so desired. Any caller temporarily lost because of a full system or as a result of a renege may decide to return after some time has passed, or never to retry (to seek help elsewhere or maybe just forget the problem completely).

Since the IRS system permits waiting but does have losses, it is a combined loss/delay system. Hence, neither of the classical Erlang $B$ or $C$ formulas (see Cooper, 1981) is applicable, since neither permits both loss and delay. Thus an alternative form of analysis was needed. Since service in the midst of congestion is the major problem, there were some simplifying assumptions made as a result of the heavy traffic. The resultant model was a bivariate Markov chain of moderate size, solved fairly routinely as a system of simultaneous linear equations, when the parameters are known.

However, it is not a simple matter to estimate system parameters from actual data. The primary difficulty is separating fresh or virgin callers from those who are retrying, since each operator has generally no way of distinguishing one from the other. But the model requires, as input, estimates for both the rate $\lambda$ of new entries and, most importantly, the critical retrial probability $\alpha$. The object of this paper then is to present an effective moment-type procedure for estimating $\lambda$ and $\alpha$. (See Appendix 1 for a complete listing of our notation.)

It is important to note that the results which follow do not depend on the queueing model, but are rather derived from stochastic balance arguments. These estimates of $\lambda$ and $\alpha$ are needed not only as input to the HHS queueing model, but also for making important IRS resource allocation decisions, and, most importantly, for estimates of customer demand, which, in turn, determines the following year's budget allocation from Congress.

3. Estimation issues

The experience of the Internal Revenue Service has clearly shown that the total number of calls into their toll-free taxpayer telephone network greatly overestimates the actual number of distinct callers using the system. In the past, IRS has measured demand by adding one-third of both overflow and abandoned call totals to the number of calls processed. But it was realized that this approach was based on unsubstantiated subjective estimates, and a study thus was begun to find a more accurate way for estimating true demand.

Data were collected by their Taxpayer Service units and transmitted to the project team for each of five key regional sites: Atlanta, GA; Houston, TX; Indianapolis, IN; Jacksonville, FL; and Richmond, VA. The major portion of the database was constructed by information from pairs of adjacent weeks in each of four vital seasonal periods for all sites. This gave a total of 40 weeks of data, with each day's statistics aggregated into 9 hourly periods, giving figures for total calls received, processed, blocked, and abandoned. To these were added a few additional observations which were already on hand and thought to provide a strong enhancement to the others. Almost all the data were in acceptable form, and in the end, very few were found faulty.

Note our assumption that decisions to retry are IID across all applicable customers in each hourly period, but with change permitted from one hour to the next. This hourly variability is matched to data availability, and the independence of $\alpha$ from prior caller history is defended in great detail in HHS. The analysis there shows, for example, that past waiting may either decrease or increase one's desire to recall, and that a periodically changing constant value is totally defensible.

The analysis was then carried out in two major steps, using the aforementioned data set. The first was a detailed sequence of calculations for estimating a daily retrial percentage estimate for each of the days (by location) in the database. This was done using a special statistical procedure whose details will be provided in the next section of this paper. The daily figures were then merged by location and/or time of year to give weighted average retrial probabilities, which are to be used then in support of budget and planning activities in a more comprehensive way.
The fundamental procedure used to develop the formulas was derived from a basic consideration of queueing theory (see Gross and Harris, 1985, and HHS). The only other effort found in the recent open literature to offer a retrial probability estimate is Fredericks and Reisner (1979). We equate the customer flow into any particular location with that which goes out, and from this balancing, obtain moment-like estimates for both the retrial rate and a corresponding estimate for the actual number of callers per hour. The routine recognizes the time dependent pattern of daily demand and explicitly takes into account. Then a total day's figure is calculated as an average of each hour weighted by its relative call activity.

4. Technical details

In order to set the stage for the technical discussion and result summary to follow, we present the notation to be used in the model and its accompanying analysis. We use \( a \) and \( f \) for the rates of abandonment and blocking, respectively. The letters \( \alpha \) and \( \lambda \) represent the retrial probability and rate of first-time incoming calls, respectively, while \( \lambda' \) will be the demand which actually enters the system (i.e., both completions and abandonments). Subscripts will be used to indicate that a particular number is appropriate for a specific period. For example, \( \alpha_t \) would be the retrial probability for hour \( t \), \( f_{t+1} \) the overflow rate for period \( t + 1 \), etc.

We begin from the fundamental balance equation for the flows into and out of the telephone system, namely,

\[
\text{offered load} = \lambda + (f + a)\alpha = f + \lambda' = \text{overflow + acceptance rates.} \quad (1)
\]

Note that this also means that the true arrival rate \( \lambda \) can be found as the observed entry rate \( \lambda' - a \) plus the frequency with which blocked and abandoned calls reenter, since

\[
\lambda = (\lambda' - a) + (f + a)(1 - \alpha). \quad (1a)
\]

Our concern then is to obtain estimates from data for both \( \lambda \) and \( \alpha \) for each hour of any day, and then to combine these to get daily and weekly average figures. (The method is not for use beginning from whole days.) Note that a reasonably unequivocal estimate of either \( \lambda \) or \( \alpha \) determines the other since \( f, a \) and \( \lambda' \) are known.

Now the key observation is that the flow equation (1) implies (using a simple difference calculus argument) that any changes in overflows and acceptances \( (f + \lambda') \) from period \( t \) to the period \( (t + 1) \) (on a given day) can arise from changes in \( \lambda \) and \( \alpha \) as (with \( \Delta x = x_{t+1} - x_t \)):

\[
\begin{align*}
\Delta t \lambda + f \Delta \alpha &+ a \Delta t + f \cdot \Delta \alpha + \Delta t \cdot f' \cdot \Delta \alpha \\
\Delta t \lambda + (f_{t+1} + a_{t+1}) \Delta \alpha &+ a_{t+1} \Delta (f + a) \\
&= \Delta t (f + \lambda').
\end{align*}
\]

That is to say, a change in the total apparent offering to the main queue comes from increments in the original offered input and/or by a change in the return probability. As expected small increments in \( \alpha \) bring out big changes in the right-hand side.

Unfortunately, both \( \alpha \) and \( \lambda \) are unknowns in this formulation. In order to begin an iterative procedure which estimates both \( \alpha \) and \( \lambda \), we need to estimate the contribution of the \( \Delta \lambda \) term to the left-hand side of the difference equation given by (2). To do so, we first note that the average virgin caller contributes (by redialing) more than one count to the total apparent offered input (i.e., \( f + \lambda' \)). When we are able to compute the relative impact on \( (f + \lambda') \) of changes in \( \lambda \), we can then eliminate the \( \Delta \lambda \) term from (2). The remainder of this section will describe a methodology for doing this.

First, we need the expected number of total calls ultimately placed by a fresh caller. This is difficult to evaluate given the very complicated nature of the queue structure, but can be reasonably well approximated in the following fashion. Assume that each redial faces a full system with probability \( p_k \). Then, noting that each caller has a potential for opting not to reenter, we see that

\[
E[\text{total calls by one caller}] = E[\text{calls by those who ultimately connect}] + E[\text{calls by those who never get in}]
\]

\[
= \sum_{n=1}^{\infty} n(1 - p_k)(\alpha p_k)^{n-1} + \sum_{n=1}^{\infty} n(1 - \alpha)\alpha^{n-1}p_k^n = \frac{1}{1 - \alpha p_k}. \quad (3)
\]
In the analysis following, we denote the right-hand side of (3) as $S$ and its reciprocal as $R$, where $S$ must, of course, be greater than one, and $R$ a fraction between zero and one. It is precisely the estimation of $R$ which permits the elimination of $\Delta \lambda$.

But we realize that $R$ is time dependent, since the retrial probability $\alpha$ and the probability of blockage $p_K$ are. Generally, we are not going to know the appropriate hourly $R$ values since the retrials could have occurred during any of the previous periods. But we are able to formulate the estimation of $R$ as part of a two-phase iterative method, one phase over time, the other with respect to $\alpha$.

To begin estimating $R$, we note that there is limited variability in $p_K$ over any given day. So we assume in the evaluation of equation (3) that $p_K$ should be set at its average value for the day of concern. Next, we begin a series of iterations (hour to hour) by setting $\alpha$ at an initial guess. All hourly retrial probabilities are then calculated as discussed shortly, and their average compared to the approximation used in (3). If they disagree, the value of $\alpha$ in (3) is updated to the new estimate, and we repeat the procedure. (Convergence in $\alpha$ has always occurred quickly for all starting points, though we have been unable to prove that such is guaranteed.)

With $R$ known, the final equation for hourly estimation can be derived from equation (3). We see that a change $\Delta \lambda$ in the originating input rate leads to a change $S\Delta \lambda$ in the offered input rate. Hence we note that $\Delta \lambda = R(\lambda + f)$ and thus rewrite equation (2) (now free of $\lambda$) as

$$
(f_{i+1} + a_{i+1}) \Delta_i \alpha + \alpha_i \Delta_i (f + a) = (1 - R) \Delta_i (f + a') \Delta_i (f + a).
$$

As an illustration of the methodology, let us consider the data of Table 1 from Jacksonville, Florida. For this example, we begin by assuming that the average retrial rate $\alpha$ needed for $R$ is equal to 0.7. We shall, however, leave the value at 0.7 throughout the illustrative calculation since an explanation of the full iteration complicates the narrative unnecessarily. Based on the day’s specific blockage experience, we calculated a value of $R = 0.46$. (In real practice, only the first full iteration of our procedure uses $\alpha = 0.7$. Then the resultant average $\alpha$ is used for the next pass, and so on.)

We start the sequence of hourly calculations by assuming that a day’s largest rate of retrial, $\alpha$, is equal to one or equivalently that the lowest loss rate of any given day is nil. Indeed, any such initial setting for $(1 - \alpha)$ permits the algorithm to work. But $\alpha = 1$ would appear to be at least as tenable as say 0.98 or 0.95. Our previous work on the problem suggests that such high return rates for the busiest period of days during the busiest times of year seem plausible. These high return rates are a very major contributor to the congestion.

To locate the most appropriate period for $\alpha = 1$, we first solve for each of the $\{\alpha_i\}$ in terms of the first one using the iteration procedure. As a result, each $\alpha$ is now represented as a linear form in $\alpha_1$, say, $a_i\alpha_1 + b_i$, $i = 2, \ldots, 7$. Then we can show that

$$
\alpha_i = \min \left(1, \min_i \left(\frac{1 - b_i}{a_i}\right)\right).
$$

This can be verified by considering the linear forms simultaneously, with the result (5) following as the basic solution with $0 < \alpha_i \leq 1$. (See the Appendix for a proof.) Use of this formula now guarantees that the maximum of the $\{\alpha_i\}$ is one. It could also theoretically happen that one such probability could actually be driven negative. But we have never seen this happen.

For the illustrative problem now, we commence by getting $\alpha_1$, in terms of $\alpha_1$ from (4) for each hourly period. The first $\alpha$ then to be calculated is

<table>
<thead>
<tr>
<th>Hour</th>
<th>Offered Load: Overflows + Accepteds</th>
<th>Overflows + Abandons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1388</td>
<td>923</td>
</tr>
<tr>
<td>2</td>
<td>1721</td>
<td>1177</td>
</tr>
<tr>
<td>3</td>
<td>1544</td>
<td>1087</td>
</tr>
<tr>
<td>4</td>
<td>1227</td>
<td>812</td>
</tr>
<tr>
<td>5</td>
<td>1666</td>
<td>1210</td>
</tr>
<tr>
<td>6</td>
<td>1213</td>
<td>724</td>
</tr>
<tr>
<td>7</td>
<td>1675</td>
<td>1211</td>
</tr>
<tr>
<td>Totals</td>
<td>10434</td>
<td>7144</td>
</tr>
</tbody>
</table>

(daily $p_K = 0.66$)
\( \alpha_2 \), which follows as
\[
\alpha_2 = \frac{f_1 + a_1}{f_2 + a_2} \alpha_1 + \frac{(1 - R)(f_2 + \lambda^2 - f_1 - \lambda^1)}{f_2 + a_2}.
\]

This leads to
\[
\alpha_2 = \frac{923}{1177} \alpha_1 + \frac{0.46(1721 - 1388)}{1177}
\]
\[
\approx 0.784 \alpha_1 + 0.1301.
\]

Next,
\[
\alpha_3 = \frac{1177}{1087} \alpha_2 + \frac{0.46(1544 - 1721)}{1087}
\]
\[
\approx 1.082 \alpha_2 - 0.0749
\]
\[
\approx 0.848 \alpha_1 + 0.0659.
\]

Then
\[
\alpha_4 = 1.339 \alpha_3 - 0.1796 \approx 1.135 \alpha_1 - 0.0914,
\]
\[
\alpha_5 = 0.671 \alpha_4 + 0.1669 \approx 0.752 \alpha_1 + 0.1056,
\]
\[
\alpha_6 = 1.671 \alpha_5 - 0.2878 \approx 1.273 \alpha_1 - 0.1113,
\]
and
\[
\alpha_7 = 0.598 \alpha_6 + 0.1755 \approx 0.761 \alpha_1 + 0.1089.
\]

Now, from our earlier discussion, we find that
\[
\alpha_1 = \min \left( 1, \min_i \left( \frac{1 - b_i}{a_i} \right) \right) \approx 0.873.
\]

Thus
\[
\alpha_2 \approx 0.815, \quad \alpha_3 \approx 0.806, \quad \alpha_4 \approx 0.899,
\]
\[
\alpha_5 \approx 0.771, \quad \alpha_6 \approx 1, \quad \alpha_7 \approx 0.773.
\]

When these seven retrial probabilities are then combined into a daily average (each contribution weighted by relative offered load), we get \( \bar{\alpha} \approx 0.839 \).

It should be noted that the specific Jacksonville result derived here is somewhat higher than the typical outcome. But this was the first dataset we analyzed, and the numbers are presented here strictly to illustrate the method.

5. Results

The most important of our results may be briefly summarized as follows:
- the average return probability taken over all locations and time periods was approximately 0.69 with some apparent dependence on location and time of year. (Note that this is not too different from IRS's previous guess of \( \frac{2}{3} \)).
- there is an observable increase (to 0.76) in the propensity to return a call close to the main filing time in mid April, and consistent stability in the 0.60 to 0.63 range throughout the other sampled periods. (Note that the extreme density of calls in the April period brings the weighted average up to the final total of 0.69.)
- there is some variation across the four cities summarized, from a high estimate of 0.76 in Richmond to a low of 0.57 in Houston. A small portion of the differences could be directly attributed to the relative size of the April samples in each of the cities. But Houston's April probabilities tended to be lower than the national April average, though much higher than its probabilities for the other times of the year;
- as a consequence of our model, we arrive at a micro formula from equation (1a) for the adjustment of apparent demand in order to estimate true demand (as a function of observations) when the retrial rate is \( \alpha \) as
\[
\lambda = (\lambda' - \alpha) + (f + a)(1 - \alpha).
\]

That is to say, the true demand is the number of processed calls plus a calculated percentage of the sum total of abandonments and blocked calls.

Some further comments on the results follow.

Recall that we were estimating the fraction \( \alpha \) of blocked callers and abandonments who decide to replace their calls in lieu of declining to retry. We have used the computer-based techniques pre-

Table 2
Group return probability estimates

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>0.69</td>
</tr>
<tr>
<td>By city</td>
<td></td>
</tr>
<tr>
<td>Atlanta, Georgia</td>
<td>0.66</td>
</tr>
<tr>
<td>Houston, Texas</td>
<td>0.57</td>
</tr>
<tr>
<td>Jacksonville, Florida</td>
<td>0.73</td>
</tr>
<tr>
<td>Indianapolis, Indiana</td>
<td>*</td>
</tr>
<tr>
<td>Richmond, Virginia</td>
<td>0.76</td>
</tr>
<tr>
<td>By time of year</td>
<td></td>
</tr>
<tr>
<td>January and February</td>
<td>0.63</td>
</tr>
<tr>
<td>Early through mid April</td>
<td>0.76</td>
</tr>
<tr>
<td>Mid through late May</td>
<td>0.61</td>
</tr>
<tr>
<td>October</td>
<td>0.60</td>
</tr>
</tbody>
</table>

* Sample thought to be too small for a meaningful result.
sented earlier to estimate $\alpha$ from our database for all seemingly important groups and subgroups of calling populations. All the data were merged, and a figure derived as a weighted average taken for the entire nation, across all sites, days of the week, and time of year. Then individual analyses were done by site and time of year.

It is interesting then to note the pattern of the estimates. The national figure is 0.69, with some important deviations from this norm. First, the return rate rises by about 25% from a very stable value around 0.61 or 0.62 during the year to a peak of approximately 0.76 just prior to the major filing milestone of April 15. This change actually appears to begin around late March and would seem to be totally as expected. On the other hand, there also appears to be some geographic variation. There is a fairly large difference between the high and low city values, with the range equal to approximately 28% of the mean. This is difficult to explain fully, but we know that a portion of the difference comes from the relative contribution each city database received from its April data. Of course, we realize that current caller behavior is very much a function of the past history of the specific center's record and its ability to satisfy prior demand. We do not know enough, for example, about Houston to say that the cause of the low retrial rates is rooted in people's expectations, but this is certainly possible.

Note that the effect of a misestimate of true demand in Houston is not as great as might be expected. Suppose the retrial probability were actually 0.50 for a non-April day in Houston, and we had used 0.60 instead. Then the ratio of estimated $\lambda$ values under the two alternatives (say, $\lambda_{0.5}$ and $\lambda_{0.6}$) would be

$$\frac{\lambda_{0.5}}{\lambda_{0.6}} = \frac{(\lambda' - a) + (f + a)(1 - 0.5)}{(\lambda' - a) + (f + a)(1 - 0.6)}.$$  

Of course, the values of the ratio depends on daily specifics. To illustrate, let us take some real numbers from Houston for January 28, 1981. We find that $\lambda' = 7386$, $f = 522$, and $a = 83$. Thus

$$\frac{\lambda_{0.5}}{\lambda_{0.6}} = \frac{7605.5}{7545.0} = 1.008,$$

or only a 0.8% error. Thus we believe that the variation in city retrial rates is not a very serious problem.

6. Concluding remarks

In addition to being able to improve estimation of true demand for personnel-planning purposes, these results will permit the completion of the detailed queueing-theoretic model of IRS sites documented in HHS, since now all necessary parameters are known. Thus both staffing and trunkline configurations can be made using analytic queueing results.

In light of all the results documented here, we have made the following recommendations to IRS:

1. It would be most useful for the continued evaluation of the system to reassess the data from key sites periodically. Whenever changes in the data are observed, the analysis can easily be redone. Thus management can regularly assess the effects of alterations in the types of services provided.

2. Continue collecting data for the key filing periods and expand the scope to include additional centers, particularly in other regions of the nation.

3. Make sure that all of our computer programs are well understood and made operational at the Service.

4. If at all possible, do some direct caller sampling to determine whether the model estimates are truly on target.

Appendix 1

**Notation**

- $\lambda = \text{average number of distinct callers attempting to enter the system per unit time}$
- $\lambda' = \text{average number of callers accepted per unit time}$
- $f = \text{rate of calls unable to enter the system because no lines are available}$
- $a = \text{number of accepted callers leaving per unit time prior to the commencement of service}$
- $\alpha = \text{probability of retrial}$

The system of equations at question is

$$\alpha_i = a_i \alpha_1 + b_i \quad (i = 2, \ldots, N).$$
Define \( z_i = 1 - \alpha_i \). Then
\[
1 - z_i = a_i \alpha_i + b_i
\]
or
\[
a_i \alpha_i + z_i = 1 - b_i \quad (i = 2, \ldots, N; \quad 0 \leq \alpha_i \leq 1, 0 \leq z_i \leq 1).
\]
(6)

For the moment, ignore the upper bounds of 1; then our problem is to find nonnegative numbers \((\alpha_1, z_2, z_3, \ldots, z_N)\) to solve (6). We see immediately that an initial basic solution is
\[
(\alpha_1, z_2, \ldots, z_N) = (0, 1 - b_2, 1 - b_3, \ldots, 1 - b_N).
\]
Now bring \( \alpha_i \) into the basis. Then one \( z_i = 0 \) (say, \( i = j \)), and
\[
\alpha_i = \min \left( \frac{1 - b_j}{\alpha_j} \right) = \frac{1 - b_j}{\alpha_j}.
\]
But \( z_i = 0 \) implies \( \alpha_j = 1 \) and all other \( \alpha_i < 1 \). If, however, this minimum is greater than 1, make \( \alpha_1 = 1 \) and \( z_i = 0 \), and we see that \( z_i > 0 \) for all \( i \neq j \). In either case, \( \max(\alpha_i) \) is still 1 and all constraints are satisfied.

References


