STOCHASTIC MODELS FOR SERVICE EVALUATION SYSTEMS

A. DESCLoux
Bell Telephone Laboratories, Inc., Holmdel, New Jersey, U. S. A.

ABSTRACT

Service evaluation is performed in order to determine the telephone network performance as seen by the users. It is therefore restricted to the short intervals when customers attempt to place calls and is thus only concerned with items such as switching-network blocking, trunk busy-conditions, "don't answer", vacant codes, switching and dialing errors, etc.

The present paper describes a partially mechanized service evaluation system with emphasis on the traffic flow and provides models that have been developed to even-out the evaluators' work loads and to improve the likelihood of evaluating a sufficiently large number of calls from each of several categories.

The appropriate models involve single-server links which may be momentarily unavailable to an incoming request because it is either already being used or because of evaluator busy-conditions. An essential feature of these systems is that the input can be monitored at will.

1. INTRODUCTION

The purpose of service evaluation is to gather information on network performance as seen by telephone users and to determine how efficiently the equipment is being used. Pertinent records therefore cover the intervals from those instants when either receivers are off-hook or trunks are seized to the times when (calls) are either answered by the called parties or fail for a variety of reasons such as switching-network blocking, trunk busy-conditions, "don't answer", vacant codes, switching or dialing errors, etc. (Other relevant indicators such as, for instance, the dial-tone speed and the transmission quality are recorded by other means.)

In each central office the facilities are grouped by type into entities and the number of calls to be observed monthly is specified for each of them. The sampling plan involves further stratifications and ultimately prescribes how many calls of each type should be observed during each 15-minute interval; furthermore, if at the end of a 15-minute interval a quota has not been met as planned, the (cumulative) deficiency is added (under restrictions stated below) to the nominal quota of the next 15-minute interval. All these operations are carried out under the direction and control of a minicomputer located at the service evaluation bureaus.

Each evaluation bureau handles calls from several central offices and is connected to each of them by a single data link (Signal Converter/Allotter-SC/A). The model investigated here deals with the case where the minicomputer keeps the "gate" open so long as quotas have not been filled: As calls originate, the lines and trunks under observation are connected to the data link (if free) and are observed if an evaluator is available. From a traffic point-of-view the data link can then be regarded as a single-server loss-system with additional blocking due to the possible lack of a free evaluator.

Under these conditions and all the assumptions needed to assure that the traffic flow over each link is Markovian, we first determine, for time intervals of arbitrary length, the joint distribution of the number of calls that are evaluated for each of the entities competing for a given link. This distribution, obtained without quota restriction, is then used to determine algorithmically the probability that the various entities meet their respective quotas over periods (bands) of arbitrary length.

The present model allows us to evaluate the effect of competition among the various entities in terms of their relative loads as well as the impact of evaluator blocking and attendant link-usage losses. It also makes it possible to determine whether or not the 15-minute quota patterns can be chosen in such a way as to even-out the evaluator's work loads as well as to appreciably improve the chance of meeting the various quotas. The answer to these last two questions are respectively mildly positive and strongly negative. It is shown, in particular, that small load variations have a strong impact on the system's behavior: quotas determined from even moderately over-estimated loads cannot be met as a rule and pattern adjustment prove then to be a useless procedure.

The computations of the basic formulas of this paper make extensive use of hypergeometric and related classical functions which have to be evaluated over ranges where numerical stability becomes a problem. Means to circumvent these difficulties are briefly discussed.

2. THE SAMPLING PLAN

Service evaluation is performed daily from 7 a.m. to 10 p.m., yet only a very small portion of the network can be observed at any one time and to guard against systematic biases the lines, trunks and positions which come under scrutiny are periodically changed. In each central office the facilities to be observed are grouped into entities such as groups of dial lines or sets of incoming trunks. For each entity the number of calls to be observed monthly is prescribed and the evaluators attempt (jointly) to meet these quotas. The sampling plan involves further stratifications: the daily observation period is split into seven bands: a 3-hour band from 7 a.m. to 10 a.m. followed by six 2-hour bands which cover the 10 a.m. - 10 p.m. interval. So called morning, afternoon and evening sessions extend, respectively, from 7 a.m. to 10 p.m., from 12 p.m. to 6 p.m. and from 6 p.m. to 10 p.m. For each type of entity, workday, Saturday and Sunday load estimates are obtained for every band from peg-counts gathered yearly for a month. The monthly quotas are then spread among the bands (workdays, Saturdays and Sundays, separately) in direct proportion to the call volumes. Finally, each band is subdivided into 15-minute intervals called epochs, and the minicomputer is told how many calls should be observed during each of these subintervals. All quota assignments (patterns) are allowed and this flexibility was introduced in the hope of evening out the evaluator work-loads and to improve the chance of meeting the quotas.

The minicomputer keeps track of the number of observations that have been made and if at the end of an epoch a quota has not been met as planned, the (cumulative) deficiency is added to the nominal quota of the next epoch. Carry-over is allowed between bands of a given session, but, if a certain band quota is not met at the end of a work-day session, that particular band-deficiency is added to the quota of the first 15-minute interval of the same band of the next work-day. Saturdays' and Sundays' deficiencies are handled in the same manner.

Each SC/A has 16 ports and can therefore handle up to 16 entities. The SC/A are directed by the minicomputer to submit calls of the desired type to the evaluators and all the entities homing on the same SC/A and which have unmet quotas, compete for the link to the Service Evaluation Center (SEC). The calls which do have access to the SC/A link are observed provided, of course, an evaluator is available which need not be the case. Actually, under normal loading conditions, evaluator-blocking can be expected to be quite high, 80 percent or more.
3. MODEL FOR THE SC/A LINK

As noted earlier, the SC/A link can be regarded as a single-server, loss-systems with input controlled by the mini-computer: As calls are generated, the lines and trunks of entities which have not met their quotas are bridged to the SEC if the SC/A is free and they are observed if, in addition, an evaluator is immediately available. If either of these two conditions is not met, the calls cannot be observed, and are lost as far as service evaluation is concerned. However, all the calls which succeed in getting the SC/A link but fail to get an evaluator, do not disappear instantly, but hold the SC/A link for a brief time interval called dead time. Although relatively short, these intervals have a cumulative effect which is not negligible since, under prevailing circumstances, blocking solely due to evaluator-busy conditions is quite high by design.

The computations required to evaluate the system are involved even under the restrictive assumptions that we now proceed to state.

We shall investigate a single channel with Poisson input of intensity \( \lambda \) and negative-exponential evaluation time with mean \( 1/\gamma \). In addition, we shall also assume that the dead times are negative exponential with mean \( 1/\gamma \) and that the probability, \( \theta \), that all the evaluators are busy is constant and known.

Under these conditions, we determine the equilibrium distribution \( P_c(t) \) of the number of calls observed during a time interval (epoch) of arbitrary length \( t \). This is done without taking the quotas into account, i.e., we disregard at first the restrictions placed on the maximum number of calls to be observed in any given interval. Thus far the input to the SC/A link is the joint input of all the entities with quota requirements in a given epoch. Let \( n \) be the traffic intensity of the \( i \)-th entity. When \( m \) entities have access to an SC/A, any number of them may not have to be sampled during a given epoch and, therefore, the distribution \( P_c(t) \) must be evaluated, in general, for \( 2^m-1 \) values of the parameter \( a, \) viz. \( a_1a_2 + + a_m \)

where \( i_1, \ldots, i_k \) are any \( k(=1, 2, \ldots, m) \) integers satisfying the condition \( 1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq m \). Whenever entities \( i_1, \ldots, i_k \) must be sampled during an epoch, the distribution \( P_{c i_1i_2} \) of the number of calls observed from entity \( i_j \) over \( (0,t) \) can be retrieved by means of the formula

\[
P_{c i_1i_2}(n) = \sum_{k=0}^{n} \binom{n}{k} \frac{a_{i_k}}{\gamma} \cdot \left(1 - \frac{a_{i_k}}{\gamma}\right)^k \cdot \left(1 - \frac{a_{i_k}}{\gamma}\right)^{-n-k},
\]

where \( i_1, \ldots, i_k \) is the conditional probability that an observed call is from entity \( i_j \), given that the competing entities are \( i_1, \ldots, i_k \).

It is sometimes reasonable to set \( a_i = 1 \) equal to the relative load intensity of entity \( i \). This is the case, in particular, when all the calls sampled via the same SC/A are observed by the same evaluator. In practice, however, this is the exception rather than the rule.

Of interest here are the total numbers of calls observed from the various entities at the end of successive epochs. Jointly, these numbers constitute a nonstationary Markov vector-process with state-dependent transition functions and boundary conditions induced by the quotas.

More specifically, let \( P_c(n_1, \ldots, n_m) \) be the probability that \( n_i \) calls from entity \( i \) \((i=1, \ldots, m)\) are observed over \( k \) consecutive epochs and let \( q_i \) be the number of calls that one should observe by the end of the \( (k+1)\)-st epoch.

Let \( e(i, k) \equiv (1-\delta_{ik}) \) where \( \delta_{ik} \) is the Kronecker symbol.

Then, with \( d_i = n_i - q_i \) and \( s \equiv E_d_i \), we have

\[
P_k(n_1, \ldots, n_m) = \sum_{c_1} \cdots \sum_{c_m} p_k(r_1, \ldots, r_m)
\]

where \( c_1 = (r_1, n_1) \) if \( n_1 < q_1 \) or \( n_1 = q_1 = r_1 \), \( c_2 = d_2 \), \( d_2, \ldots, \ldots \) if \( n_1 > q_1 > r_1 , \) \( v(c_i) = (p_k/E_p_k(r_1, q_1)) \) and \( a = E_q_a(r_1, q_1) > 0 \).

The preceding recursion formula allows us to compute the probabilities of meeting the quotas at the end of any sequence of consecutive epochs. This process turns out, however, to require a considerable amount of computer time. To keep the computations within reasonable bounds the dimensionality of the model has to be low. Consequently, all our programs pertain to the case where only two entities have access to the SC/A. But this is not as restrictive as it may seem. Indeed we may assume that all but one entity are pooled together. The model may then be used to evaluate what happens to the segregated entity as it competes with the others for the SC/A link. Alternatively, one could assume that all but one or two entities are mixed with what thus far has been the "dead time" traffic. This can be done by means of a suitable choice of \( \theta \) and of a new "dead time" distribution. Note that this reinterpretation allows the study of entities with unequal average evaluation times since the average "dead time" can be chosen at will. This arbitrariness allows further flexibility as indicated next.

As a rule the dead times are relatively short (as compared to the average evaluation time) and the assumption that \( \theta \) is constant is probably not always realistic. Busy conditions tend to persist and the calls which arrive shortly after the conclusion of a dead time are likely to experience evaluator-blocking in excess of \( \theta \). This is particularly reasonable to assume since the evaluators do not always become immediately available at the conclusion of an observation. To compensate for this effect one could artificially extend the length of the dead time and add the requirement that the busy period be a multiple of the evaluator busy-period. With this interpretation of the dead time, the model can now also be used to evaluate the effect of adding or removing evaluators since such changes directly affect the length of the busy period.

In the analysis, \( \theta \) is defined as the conditional probability that a call cannot be observed for lack of an evaluator, given that the SC/A link is free. Under the condition that the SC/A link is not being used to make an observation, the probability \( \theta^* \) that a call cannot be observed because of dead time and/or evaluator busy-conditions is given by the formula (we now set \( s=1 \))

\[
\theta^* = 0 \ (a/y) + a(1-a) \ (a/y) \ 0 \ \gamma (a/\gamma) + 0 \ \gamma (a/\gamma) \ \gamma (a/\gamma)
\]

with equality iff \( \gamma = \frac{1}{\gamma} \).

By means of this formula, we can evaluate the proportion of time when the SC/A link is used to evaluate a call (referred to below as the SC/A capacity) is equal to

\[
a(1-\delta)/(a(1-a)) = \frac{a\gamma(1-a)}{\gamma + a\gamma + a\gamma(1-a)}
\]

For given values of \( \theta \) and \( \gamma \), this expression increases monotonically as a increases and the maximum capacity of the SC/A link is therefore

\[
\gamma(1-a)/(\gamma + a\gamma + a\gamma(1-a))
\]

Let \( c \) stand for the SC/A capacity. Then, by (1), we have

\[
a = c \gamma/(1-c) \gamma (1-a) - c \gamma
\]

and, for any prescribed \( \gamma \), the value \( c \) of \( \theta \) for which the denominator in (3) vanishes, corresponds to the maximum SC/A capacity:

\[
\gamma = (1-c)/(1-a) - 1/(1-a)
\]
Stated otherwise, for a given $y$, the capacity $c$ can be reached if and only if $0 < x/y$.

Finally, the SC/A link occupancy, $w$, is given by

$$w = \left\{ \begin{array}{ll} 1 & \text{for } x = 0, \text{ and otherwise} \\ \frac{q_1(1-E_{e_1})}{(1-x)(1+y)} = q_{(1-x)} & \text{for } x > 0 \end{array} \right. \text{for } y \neq 0$$

Clearly $w = 0$ if and only if $y = 0$ and/or $y$ is infinite.

The model considered here has one shortcoming: the input of any entity which meets its quota during a given epoch is not instantly set equal to zero but remains unchanged until the end of that epoch. One immediate consequence of this undesirable feature is to subdivide each epoch into shorter subintervals of equal length and to assign each full epoch quota to the first subinterval (unrestricted competition between subintervals remains of course allowed). However, as the subintervals become small (as they should if one wants to closely simulate instant input-cutting when quotas are met) the number of calls observed during consecutive subintervals become increasingly dependent. As shown elsewhere, the model can, however, be modified to take these dependences into account. But the computation cost is then greatly increased.

4.0 SYSTEM'S PERFORMANCE

From our point of view, the evaluation of the system's performance involves four distinct components: (i) the probability of meeting the quotas or given fractions thereof, (ii) the individual evaluator work-loads and to a lesser degree (iii) the SC/A-link capacity. These indicators, which are considered in Sections 4.1-4.3, should all be relatively high. However, they are interdependent and some compromise must be reached.

4.1 Meeting Quotas

The epoch quotas are set up in accordance with calling-rate estimates and do not reflect the seasonal variations of the loads. Furthermore, they are office wide and thus do not anticipate directly to the particular groups of incoming trunks and dial lines being observed. Hence, over any period of time, the load expected from a given entity may be substantially lower or higher, or just about equal to the load actually submitted by that entity. These variations will often determine whether some of the quotas are reachable or not: overestimates of calling rates which are bound to happen will produce unreasonably high quotas which it is clearly unreasonable to expect to fill.

As a rule, epoch quotas are used to even-out the evaluator loads over the sessions. However, exceptions are made for entities with relatively light calling rate since, as one may suspect, their likelihood of failing to meet their quotas tends to be higher. In these instances, the first epoch quota of a given hour (or band) may be set equal to the full hourly (or band) quota, a procedure referred to as front-loading. If one should, however, realize that front-loading can, at best, be only marginally effective because any entity which fails with some consistency to meet its quota is automatically front-loaded: Indeed, the deficiencies being carried over quickly make the nominal quotas inoperative anyhow. And, any entity which cannot produce reliably a given number of calls is not likely to be helped if allowed to submit an even greater number of calls.

Figures 1 and 2 show the cumulative deficiencies of two competing entities. Here we can observe the typical behavior of entities which cannot meet their quotas on time. The cumulative deficiencies drop quickly to 0 only after the respective entity has met its quota. One may then observe that the deficiencies being carried over quickly make the nominal quotas inoperative anyhow. The rippling effect of front-loading notwithstanding.

We now focus attention on entity 2. Let $q_{1}, q_{2}, \ldots$, be its successive epoch-quotas and let $e_{n}$ be the random number of evaluated calls made during the nth epoch. In Figure 1 the plotted expected deficiencies, $d_{n}$, at the end of each epoch $n$, were computed by means of the formula:

$$d_{n} = \sum_{i=1}^{n} (1-E_{e_i})$$

restricted to $n = 1$.

where $q_{1} = 5, E_{e_1} = 16$ and $q_{1} = 0$ otherwise.

To evaluate the effect of front-loading, it is useful to define the deficiencies in a different way:

$$d_{n} = \sum_{i=1}^{n} (q_{i}-E_{e_i})$$

with this definition, the deficiencies are computed with respect to an "average quota." This corresponds to the pattern 5 5 5 16 0 0 0 0 in Figure 2 was obtained in this manner.

In Figures 1 and 2 the expected deficiencies at the end of the fourth, eighth and twelfth epochs are of course equal. In Figure 2, however, gives a better description of the impact of front-loading: It allows the entity to get ahead but not to catch up. Indeed, at the end of epoch 1, 2 and 3, the expected deficiencies are actually negative. With the parameter values of Figures 1 and 2, we infer, from Figure 5, that the probability that entity 2 meets 90 percent or more of its quota is slightly larger than .7. The corresponding probability for entity 1 is much smaller than .7 and cannot be accurately read. In Figure 5, the position blocking is set at .8 and the dead time at six seconds. The strong effect of the dead time can be judged by comparison with the case where it is equal to 0 (Figure 3). Under this condition we see that the probabilities that entities 1 and 2 each meet at least 95 percent of their quotas are now, respectively, in excess of .7 and .95.

Another way to evaluate the impact of the dead time, $D$, is in terms of the number of additional calls needed to compensate for its presence. When $D = 0$ and $\theta = 0$, entities 1 and 2 must generate respectively, on the average, $45 (=9/\theta)$ and $37.5 (=7.5/\theta)$ calls per epoch in order to meet at least 95 percent of their quotas with probability .95 (see Figure 3). In 6 seconds these numbers are inflated to 77.5 and 60, respectively (see Figure 4). The magnitude of the differences of these values should serve as a warning that the dead times should be strictly controlled.

Further study of Figures 3-5 reveals that the probabilities of meeting various proportions of the quotas are very sensitive to load variations: the ranges .7 < $P_{i}$ < .95, i = 1, 2, corresponds to rather narrow ranges of $A_{1}(1-0)$. The effect of competition between entities is also illustrated by these charts.

It should also be noted that for any set of quotas the largest observable percentage of the quotas cannot exceed a certain value $R_{\text{max}}$ (<1) which is approached, but never reached, as the loads tend to $\infty$. In the case of a single entity with constant epoch quotas $Q$, it is readily seen by means of (2) that $R_{\text{max}}$ = $\frac{Q}{(1-0)(1+y)}$, where $t$ is the average evaluation time.

4.2 Evaluator Work-Loads

In principle, the number of evaluators can be varied at will to control their respective work-loads. But in fact high individual work-loads cannot be achieved unless the system operates in a region of relative instability where small changes in dead times and position blocking may have a crucial impact on the system's ability to meet quotas. This can be seen from Figure 6, where the percentage of meeting at least 95 percent of the quotas is expressed in terms of the position blocking and of the average dead time, $D$, in multiple of the average evaluation time. (For meaningful comparison the loads $A_{1}, A_{2}$, offered in each instance are such that $A_{1}(1-0)$ and $A_{2}(1-0)$ remain constant.)

The precipitous dip for position blocking in excess of .8 illustrates the pitfall of high individual work-loads.

4.3 SC/A Link Capacity

As defined above, the SC/A link capacity is equal to the expected value of the proportion of time when the link is used to evaluate calls. Figures 7 display maximum capacities (which can only be approached if the loads increase beyond all bounds) in terms of the average dead time and position blocking. But because of load limitations, the maximum capacities are hard to approach as can be seen from Figure 6 where the average dead time is fixed and the total load is allowed to vary.
5. CONCLUSIONS

All the computations and examples clearly show that the system studied here reveals the typical volatility of the familiar random walk processes and that even evaluator loads and high probability of meeting quotas are, at best, difficult to achieve systematically by means of the proposed quota controls. A way to achieve greater control would probably have to involve counters to keep track of the proportion of calls which have been evaluated and direct that observations be attempted so long as that proportion falls below a certain value. Such arrangements would have the following advantages: (i) the samples would be proportional to the loads generated by the various entities, (ii) priority would always be given to the entities that are falling behind and (iii) the sampling rate could be dynamically adjusted so as to maintain adequate evaluator work-loads.

REFERENCES


APPENDIX A

Distribution of the Number of Calls Carried by a Channel Without Waiting Position

Consider a single channel (server) with Poisson input of intensity $\lambda$ and negative-exponential service time with mean $1/\mu$. We investigate here the case where calls which arrive when the channel is busy are not instantly dismissed but block the access to the system for a time interval which is assumed to be negative-exponential with mean $1/\gamma$. We now determine the distribution $\{P_n(t), n=0,1,2,\ldots\}$ of the number of calls that are served during a time interval of length $t$.

Let $P_0(t)$ be the conditional probability that $n$ calls are carried by the channel during $(0,t)$ given that the channel is free at time 0 ($n=0,1,2,\ldots$). These probabilities satisfy the following relations:

$$P_0(t) = e^{-\lambda t} + \alpha(1-\beta)S(t-u)^{\gamma}$$

$$P_1(t) = \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

$$P_2(t) = \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

$$P_n(t) = \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

We define the following probabilities:

(i) $P_0(t)$, the probability that the server is free at time 0,

(ii) $P_1(t)$, the probability that a call is being served at time 0, and

(iii) $P_2(t)$, the probability that the server is free at time 0 but cannot accept a request because access is momentarily blocked.

Let us now assume that the time origin is an equilibrium instant. A simple renewal theoretic argument shows that

$$P_1(t) = e^{-\lambda t} + \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

with $P_1(t) = \beta(s+\gamma)^{-1}P_0(t)$ and $P_0(t) = \gamma(s+\gamma)^{-1}P_0(t)$.

We then have, by means of (A1-A5) and upon setting $s$ equal to 1:

$$P_n(t) = \gamma(s+\gamma)^{-1}P_{n-1}(t) + P_1(t)P_{n-2}(t)$$

Let us now assume that the time origin is an equilibrium instant. A simple renewal theoretic argument shows that

$$P_1(t) = e^{-\lambda t} + \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

$$P_2(t) = \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

$$P_n(t) = \alpha\beta \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du dv$$

We then have, by means of (A1-A5) and upon setting $s$ equal to 1:

$$P_n(t) = \gamma(s+\gamma)^{-1}P_{n-1}(t) + P_1(t)P_{n-2}(t)$$

and

$$P_n(t) = A_n(s+\gamma)^{-1} + P_1(t)P_{n-2}(t)$$

where

$$A_n = \gamma(s+\gamma)^{-1}$$

and

$$H_n(s) = \frac{\gamma(s+\gamma)^{-1}}{(s+\gamma)^{n+1}}$$

A series of algebraic manipulations then yield:

$$J_n(s) = \sum_{j=0}^{n+1} \left[ (-1)^{n-j} \frac{\gamma(s+\gamma)^{-1}}{(s+\gamma)^{n-j}} \right] J_{n-j}(s)$$

where

$$J_n(s) = \frac{\gamma(s+\gamma)^{-1}}{(s+\gamma)^{n+1}}$$

In view of (A6), the inverse of $P_n$ can be expressed in terms of the inverses $J_{n-j}$ for $j = 0,1,2,\ldots$. Let $K_n$ be the inverse of $P_n$. Then

$$K_n(t) = \frac{1}{(n-1)!} \int_0^t e^{-\lambda(u-v)}P_0(t-u-v)du$$

and

$$K_n(t) = \int_0^t \frac{1}{(nt)^2} \int_0^t \frac{1}{u^2} e^{-\lambda(u-v)}P_0(t-u-v)du$$

We define the following probabilities:

(i) $P_0$, the probability that the server is free at time 0,
\[
\frac{d^v}{dy^v} \sinh\left(\frac{\alpha y}{2}\right) \bigg|_{y=1}
= \frac{\alpha^v}{n!} \left(\frac{1}{\alpha}\right)^{2n+1} e^{-\frac{\alpha}{2}} \sum_{j=0}^{\infty} \frac{\alpha}{j!} (-1/2)^j (n-v)
\]

(A8)

In the preceding expressions \(-r_1\) and \(-r_2\) are the two roots of \(s^2 + (\alpha + \gamma)s + \beta_0 = 0\) and \(s\) and \(\Delta\) are equal to \((r_1 + r_2) / 2\) and \((r_1 - r_2) / 2\), respectively. (The derivation makes use of the relation

\[
(i)_{(n)} = \frac{d^n}{dy^n} I_{1,1}^{(\alpha + \gamma, \alpha + \delta, \gamma)}(y(\alpha + \gamma), y(\alpha + \delta), y)
\]

Substituting (A8) into (A7) we find that

\[
J_{n_1}(t) = \frac{1}{n!} \left(\frac{1}{\alpha}\right)^{2n+1} \frac{\alpha^v}{j!} \left(-1/2\right)^j (n-v) \int_0^1 e^{-t\frac{\alpha}{2}} \left[\frac{d^j}{dy^j} \sinh\left(\frac{\alpha y}{2}\right) \bigg|_{y=1}\right] dy
\]

where

\[
I_j(t, z) = \frac{1}{j!} \int_0^1 e^{-t\frac{\alpha}{2}} \left[\frac{d^j}{dy^j} \sinh\left(\frac{\alpha y}{2}\right) \bigg|_{y=1}\right] dy
\]

and \(M\) is Kummer's confluent hypergeometric function.

Let

\[
D[f(t)] = \left(\frac{d^2}{dt^2} + (1+\alpha \gamma) \frac{d}{dt} + [\gamma(1+\alpha \delta) + (\alpha - \delta)]\right) f(t)
\]

and

\[
P_n[I_j(t, y^{1/2})] = \frac{1}{2n!} \left[\frac{d^j}{dy^j} \sinh\left(\frac{\alpha y}{2}\right) \bigg|_{y=1}\right] \bigg|_{y=1}
\]

Then, since

\[
I_j(t, y^{1/2}) = \frac{1}{j!} e^{-t\frac{\alpha}{2}} M(1, j+1, \alpha y^{1/2} + 1 - t)
\]

and (cf. [1], p. 507, formula 13.4.8)

\[
\frac{d^n}{ds^n} M(1, j+1, z) = \frac{n!}{z^{n+1}} M(n+1, n+j+1, z),
\]

we have

\[
\sum_{j=0}^{\infty} \frac{1}{n!} c_{n,j} \left[\frac{d^j}{dy^j} I_{1,1}^{(\alpha + \gamma, \alpha + \delta, \gamma)}(y(\alpha + \gamma), y(\alpha + \delta), y)ight]
\]

\[
= \frac{1}{(j+1)!} \left(\frac{\alpha}{\alpha^v}\right)^j M(n+1, n+j+1, (1-r_2) t) - M(n+1, n+j+1, (1-r_2) t) \bigg|_{t=0}
\]

\[
= \frac{1}{n!} c_{n,j} \left(\frac{\alpha}{\alpha^v}\right)^j c_{n,j} \bigg|_{t=0}
\]

(A12)

\[
c_{n,j} = \left(\frac{n+1}{n+1}\right)^{1/2} c_{n,j} \bigg|_{t=0}
\]

(A13)

(For a given \(n\), equations (A11) are obtained by subjecting (A10) to an \(n\)-fold iteration of the operator \(y^{1/2} dy \) and then setting \(y = 1\).)

Clearly (A9) can now be written as follows:

\[
J_{n_1}(t) = \frac{1}{n!} \left(\frac{1}{\alpha}\right)^{2n+1} \frac{\alpha^v}{j!} \left(-1/2\right)^j (n-v) \int_0^1 \left[I_j(t, y^{1/2})\right] dy
\]

so that

\[
P_m(t) = A_m \sum_{j=0}^{\infty} \left[\gamma-j \sum_{n=0}^{m} a_n \left(m-j\right) + a_0 \right]
\]

(All)

The probabilities \(P_m(t)\), \(m = 1, 2, \ldots, n\), can then be evaluated by solving the system of linear equations (All) for \(j = 0, 1, 2, \ldots, n+1\), and then substituting the appropriate values in (All). (Note that (All) is lower-triangular.)

In order to apply the method sketched above we must of course first compute the \(M's\). A few words should be added in this connection.

For \(1 < \alpha < 80\) we have evaluated the \(M\)'s by means of their expansion in series of positive terms, (cf [1], pp. 504-505). But for \(1 < \alpha > 80\), we have made use of an expression obtained from the asymptotic formula 13.5.1 of Ref. ([1] p. 508) which is exact for integer parameter-values:

\[
M(1, j+1, z) = \frac{1}{\Gamma(i+1)} \left[\frac{1}{\Gamma(j+1)} \sum_{n=0}^{i} \frac{\alpha^n (j)_{(j+2n+1)} z^{-n}}{n!} + \frac{z^{-j}}{\Gamma(j+1)} \sum_{n=0}^{m} \frac{\alpha^n (j)_{(j+2n+1)} z^{-n}}{n!} \right]
\]
CUMULATIVE DEFICIENCIES

QUOTA 1 = 60 QUOTA 2 = 48
PATTERN 5 5 5 5/6 0 0 0
A1 (1-1) = 8 A2 (1-1) = 8

CUMULATIVE DEFICIENCIES

QUOTA 1 = 60 QUOTA 2 = 48
A1 (1-1) = 8 A2 (1-1) = 8

EPOCH LENGTH = 15 MINUTES
AVERAGE EVALUATION TIME = 30 SECONDS
A0 = AVERAGE NUMBER OF CALLS PER EPOCH
GENERATED BY ENTITY N
+ POSITION BLOCKING : 8
DEAD-TIME = 6 SECONDS

FIG. 1

PROBABILITY THAT AT LEAST 95% OF THE QUOTAS ARE MET

QUOTA 1 = 60 QUOTA 2 = 48
PATTERN 5 5 5 5/6 0 0 0

LENGTH OF THE SESSION = 3 HOURS
T = EPOCH LENGTH = 15 MINUTES
AVERAGE EVALUATION TIME = 30 SECONDS
A0 = AVERAGE NUMBER OF CALLS PER EPOCH GENERATED BY ENTITY N
Pn = PROBABILITY THAT ENTITY N MEETS AT LEAST 95% OF ITS QUOTA
θ = POSITION BLOCKING : ARBITRARY
DEAD TIME = 6 SECONDS

FIG. 3
PROBABILITY THAT AT LEAST 90% OF THE QUOTAS ARE MET

-length of the session = 3 hours

-epoch length = 15 minutes

-average evaluation time = 30 seconds

\[ A_n = \text{average number of calls per epoch generated by entity } n \]

\[ P_n = \text{probability that entity } n \text{ meets at least 90\% of its quota} \]

\[ \theta = \text{position blocking} = 0.8 \]

-dead time = 6 seconds

LENGTH OF THE SESSION = 3 HOURS

t = EPOCH LENGTH = 15 MINUTES

AVERAGE EVALUATION TIME = 30 SECONDS

\[ A_n = \text{average number of calls per epoch generated by entity } n \]

\[ P_n = \text{probability that entity } n \text{ meets at least 95\% of its quota} \]

\[ \theta = \text{position blocking} = 0.8 \]

DEAD TIME = 6 SECONDS

FIG. 4

FIG. 5

FIG. 6

ITC-9

DESCLOUX-7
FIG. 7

SC/A MAXIMUM CAPACITY LEVEL LINES

FIG. 8

SC/A-LINK CAPACITY