Redialing Policies: Optimality and Success Probabilities

E. G. Coffman, Jr. and *E. N. Gilbert*
Bell Labs, Lucent Technologies
Murray Hill, NJ 07974

Y. A. Kogan
AT&T Labs, Holmdel, NJ 07733

Abstract

Since callers encountering busy signals often want to redial, modern communication networks have been designed to provide automatic redialing. Redialing services commonly have two parameters: a maximum number n of retries and a total duration τ over which retries are to be made. Typically, retries are made at evenly spaced time intervals of length τ/n until either the call succeeds or n retries have failed. This rule has an obvious intuitive appeal; indeed, among the main results of this paper are proofs that τ/n-spacing is optimal in certain basic models of called-number behavior. However, it is easy to find situations where τ/n-spacing is *not* optimal, as the paper verifies.

All of our models assume Poisson arrivals, but different assumptions are studied for the call durations; for a given mean, these are allowed to have the relatively high-variance exponential distribution or the zero-variance distribution concentrated at a point. We approximate the probability of success for the Erlang loss model with $c > 1$ trunks, and we calculate exact probabilities of success for the $c = 1$ Erlang model and the model with one trunk and constant call durations. For the latter model, we present two intriguing conjectures, one about the optimal choice of τ when $n = 1$ and one about the optimality of the τ/n-spacing policy. In spite of their apparent simplicity, these conjectures seem difficult to resolve. Finally, we study policies that continue redialing until they succeed, balancing a short mean wait against a small mean number of retries to success.

1 Introduction

Automatic redialing is a relatively recent telephone service; see for example [3] where several products with automatic redialing are mentioned. Redialing services commonly have two parameters: a maximum number n of retries and a length τ of the time interval during which retries are made. When an initial call fails, i.e., finds a busy signal, redialing services typically make retries at evenly spaced intervals of length τ/n

until either the call succeeds or n retries have also failed. This paper proves that, although τ/n-spacing is not always optimal, it is indeed optimal in certain basic models of called-number behavior to be described in the following paragraphs. We also study policies that continue redialing until they succeed, balancing a short mean wait against a small mean number of retries to success.

Section 2 defines redialing policies and the mathematical models of telephone traffic analyzed in the remainder of the paper. All models assume Poisson arrivals, but the periods during which lines are busy are allowed to be constant or to have an exponential distribution. Constant busy periods of length T provide an interesting contrast to the highly variable exponential busy periods. A constant call duration models services giving recorded announcements such as weather reports, sports scores, movie schedules, etc.

The first two models that we analyze assume a single line/trunk and constant or exponential call durations; the third continues with exponential call durations, but assumes the Erlang loss model with $c > 1$ trunks. (The case $c = 1$ of the Erlang model is brought out separately because it leads to much simpler results.) The Erlang loss model describes situations where calls are placed through an exchange with $c > 1$ trunks, and failure to complete a call occurs only when all trunks are in use.

Our models apply exactly to a single redialer competing with ordinary customers, who simply leave if their initial dialing attempt fails. However, the models should be reasonable when only a small fraction of the customers are redialers. Problems with many competing redialers are much more difficult. Note also that we do not model directly dialing attempts that succeed in ringing the called party but fail because there is no answer.

Section 3 contains our main results on the optimality of the τ/n-spacing policy. Probabilities of success given in Section 4 for the τ/n-spacing policy have simple expressions only for a single trunk and either exponential holding times or constant holding times with $\tau \leq T$. Section 4 also supplies approximations for the probability of success in the Erlang loss model and calculates probabilities of success for the model with constant busy periods when $n = 1$, and for the model with n arbitrary, $\tau = nT$, and the T-spacing policy. The last two results lead to interesting conjectures about the optimal choice of $\tau = T$ when $n = 1$ and about the optimality of the T-spacing policy when $\tau = nT$. These conjectures look simple but apparently do not have a simple proof.

Section 5 shows that a policy with uneven spacing is appropriate when the arrival rate is very small, as may be the case for a local call to a number without much traffic. For the Erlang loss model, we study a random policy that significantly simplifies the success-probability calculation. A fixed number of retrials may be unacceptable to some dialers because failure may occur on all n trials. This motivates Section 6, where we study policies that continue redialing automatically until eventual success.

2 Models and policies

We assume that calls arrive in a Poisson stream at mean rate a. Calls have independent holding times with a common distribution and mean T. The form of the distribution will depend on the application. A convenient dimensionless parameter will be the traffic intensity

$$\rho = aT.$$

In the simplest application, dialing attempts fail because called numbers are themselves busy. Busy periods then represent typical phone calls. In the *exponential model*, we assume a single trunk and take the holding time distribution to be exponential with mean T.

We also consider the *constant model* with a single trunk and holding times of constant duration T, to contrast the high-variance exponential periods with periods of zero variance. With constant busy periods, the call that blocked the initial try at time 0 is sure to end before time T, which suggests that a good strategy might deliberately make all retries before time T, trying to succeed before any competing calls. This strategy is analyzed in the next section.

In other applications, busy periods are apt to be periods when a switching system has all trunks busy. Suppose callers dial through a switch with c trunks and that dialing failures occur only when all trunks are busy. With Poisson arrivals again, and with exponential holding times, the switch is modeled as Erlang's loss system (Riordan [9]). The *Erlang model* is a Markov chain of birth-death type with $c+1$ states representing the numbers $0, 1, \ldots,$ or c of busy trunks. A dialing attempt fails only if the state is c. State k has transition rates $P(k \to k+1) = a$ and $P(k \to k-1) = k/T$ (except that transitions $c \to c+1$ and $0 \to -1$ do not occur). The stationary state probabilities

$$p_k = \frac{\rho^k/k!}{\sum_{0 \le i \le c} \rho^i/i!}, \tag{1}$$

increase with k if $\rho > c$, and peak at $k = \rho$ if $\rho < c$. We are most interested in cases with ρ near c or larger so that calls have a high probability of being blocked. Although the exponential model is a particular case of the Erlang model with $c = 1$, it is instructive to study the exponential model separately because of its simplicity.

In general, an initial calling attempt that fails occurs at a random time t_0 during a busy period. A *redialing policy* specifies n waiting times $X_1 < \ldots < X_n \le \tau$. The j-th retry is then made at time $t_0 + X_j,\ j = 1, \ldots, n$, if retries $1, \ldots, j-1$ all fail. A basic problem studied in this paper is: Given n and τ, find an *optimal* policy, i.e., one that maximizes the probability $P(n, \tau)$ that one of the retries succeeds.

A policy with retries spaced a constant time x apart, i.e., with $X_k = kx$ for $k = 1, \ldots, n$, will be called the *x-spacing* policy. Of course, $x \le \tau/n$ is required but the ∞-spacing policy has some interest when large spacings x are acceptable. For any busy-period distribution with mean T, widely spaced retries fail independently with probability $\rho/(1+\rho)$; then the ∞-spacing policy succeeds with probability

$$P(n) = 1 - \left(\frac{\rho}{1+\rho}\right)^n. \tag{2}$$

3 Optimality of τ/n-spacing

Convexity arguments feature in the proof of the following theorem, the main result of this section.

Theorem 1. *The τ/n-spacing policy is optimal in the exponential and the Erlang loss models. It is also optimal in the constant model if $\tau \leq T$.*

Proof. Define

$$G(x) := P\{\text{line is busy at time } x | \text{line was busy at time } 0\}.$$

Consider any policy with $X_1 = x_1$ and $X_k = X_{k-1} + x_k,\ k = 1, \ldots, n$. Suppose a dialing attempt at time 0 fails and let $Q(x_1, \ldots, x_n)$ denote the conditional probability that retries at times $X_1, \ldots, X_n$ all fail.

The exponential and Erlang models will be covered first. In these models, whenever a redial fails, the state of the model is known to be c. The conditional probability of failing again, by redialing after waiting time x, is then $G(x)$. It follows that

$$Q(x_1, \ldots, x_n) = G(x_1) \cdots G(x_n). \tag{3}$$

Exponential model. Recall that busy periods have exponentially distributed durations with mean T and that calls arrive at rate a. The line is busy at time $x + dx$ if either (i) it is busy at x and there is no hang-up in $[x, x + dx]$ or (ii) it is idle at x and there is a new arrival in $[x, x + dx]$. This observation leads routinely to the differential equation

$$\frac{dG}{dx} = -\frac{1}{T}[(\rho + 1)G + 1].$$

The solution with $G(0) = 1$ is

$$G(x) = \frac{\rho + e^{-(1+\rho)x/T}}{1 + \rho}. \tag{4}$$

Since $\log G(x)$ is convex,

$$\log Q(x_1, \ldots, x_n) = \sum_{k=1}^{n} \log G(x_k) \geq n \log G(\sum_{k=1}^{n} x_k/n).$$

For a given value of $X_n = \sum_{1 \leq k \leq n} x_k = \tau$, the policy succeeds with probability

$$P(n, \tau) = 1 - Q(x_1, \ldots, x_n) \leq 1 - G^n(\tau/n). \tag{5}$$

The optimal policy then takes $x_k = \tau/n$ and achieves its upper bound in (5), which completes the proof for the exponential model.

Erlang loss model. In this model, an arbitrary policy fails with a probability $Q(x_1, \ldots, x_n)$ of the same form as in (3), where now $G(x)$ is the probability, starting from state c, of being in state c again after time x. Derivations of $G(x)$ appear in Riordan [9, p. 85] and Benes [1, p. 208] (where $G(x)$ is called the *recovery function*). To calculate $G(x)$ one must first find the zeros of

$$R(s) = \sum_{0 \leq j \leq c} \binom{s + j}{j} \frac{\rho^{c-j}}{(c - j)!}, \tag{6}$$

a polynomial in s. The zeros s_j of $R(s)$ are all real and negative. In terms of the s_j, $G(x)$ is computed from

$$G(x) = \frac{\rho^c/c!}{\sum_{0\le k\le c}\rho^k/k!} - \sum_{j=1}^{c} \frac{e^{s_j x/T}}{s_j} \prod_{i\ne j}\left(1 - \frac{1}{s_j - s_i}\right). \tag{7}$$

The function $G(x)$ is convex because the exponentially decaying terms of (7) have positive coefficients, a fact proved by Haantjes [4] and Ledermann and Reuter [7]. But, we need $\log G(x)$ to be convex in order to prove the optimality result for τ/n-spacing and hence to prove that the success probability is

$$P(n,\tau) = 1 - G^n(\tau/n). \tag{8}$$

To prove convexity of $\log G(x)$, and hence (8), write (7) as

$$G(x) = \sum_{i=0}^{c} C_i e^{-r_i x}$$

with coefficients C_i and exponent factors r_i both known to be non-negative. Differentiate $\log G(x)$ twice and get $(GG'' - (G')^2)/G^2$. Convexity of $\log G(x)$ will follow if $GG'' - (G')^2 \ge 0$, i.e., if

$$\sum_{i,j}(r_j^2 - r_i r_j)C_i C_j e^{-(r_i+r_j)x} \ge 0.$$

In this sum, the $c+1$ terms with $i=j$ all vanish. The remaining terms can be combined in pairs having the same exponential factor. Thus, the terms with $(i,j) = (a,b)$ and $(i,j) = (b,a)$ combine into

$$(r_a^2 + r_b^2 - 2r_a r_b)C_a C_b e^{-(a+b)x} = (r_a - r_b)^2 C_a C_b e^{-(a+b)x} \ge 0.$$

Then $GG'' - (G')^2 \ge 0$, $\log G(x)$ is indeed convex, (8) follows, and we have completed the proof for the Erlang loss model.

Constant model. In this model, a policy with $\tau \le T$ can fail in only $n+1$ mutually exclusive ways. One way, with probability $1 - \tau/T$, is that the original call lasts longer than time $X_n = \tau$. The k-th of the remaining n ways to fail requires the original call to end between times X_{k-1} and X_k, say at time $X_{k-1} + t$, $0 < t < x_k$, and for a new call to arrive between times $X_{k-1} + t$ and X_k. Here, we assume that the original call arrived during steady state, so the hang-up time $X_{k-1} + t$ is uniformly distributed with probability density $1/T$. Given t, the k-th failure occurs with conditional probability $1 - e^{-a(x_k - t)}$. An integral over $0 \le t \le x_k$ removes the conditioning on t and gives

$$q(x_k) = \frac{a x_k + e^{-a x_k} - 1}{\rho} \tag{9}$$

for the probability of a k-th failure. Then failure occurs with probability

$$Q(x_1, \ldots, x_n) = q(x_1) + \cdots + q(x_n) + 1 - \tau/T, \tag{10}$$

where $\sum_{1\le k\le n} x_k = X_n = \tau$. Since $q''(x) > 0$, a convexity argument shows that the choice $x_k = \tau/n$ is optimal. This completes the proof of the theorem. $\square$

4 Probabilities of success

Probabilities of success for the τ/n-spacing policy have simple expressions only for the exponential model and for the model with constant busy periods when $\tau \le T$. In the exponential case, the success probability is, by (4) and (5),

$$P(n,\tau) = 1 - \left(\frac{\rho + e^{-(1+\rho)\tau/(nT)}}{1+\rho}.\right)^n \tag{11}$$

Note that $P(n,\tau)$ in (11) is an increasing function of τ. Since $P(n,\tau)$ approaches $P(n)$ in (2) for large τ, the ∞-spacing policy would be optimal if long waits between redials were allowed. Comparing (2) and (11), one sees that little is gained by taking

$$\tau/n \gg [T/(1+\rho)] \ln \rho.$$

By (9) and (10), the optimal policy in the constant model with $\tau \le T$ has success probability

$$\begin{aligned} P(\tau,n) &= 1 - Q(\tau,\ldots,\tau) \\ &= \frac{\tau}{T} - nq\left(\frac{\tau}{n}\right) = \frac{a\tau - n(1-e^{-a\tau/n})}{\rho}. \end{aligned} \tag{12}$$

Note that $P(\tau,n)$ is again an increasing function of τ; $\tau = T$ is a best choice if any τ in $0 < \tau \le T$ is allowed.

For the Erlang loss model, we illustrate in Figure 1 a family of curves of $G(x)$ vs x/T for $c = 20$ trunks and $\rho = 2, 16, 30$. As x becomes large, these curves flatten out to approach asymptotes that represent the stationary probability p_c of (1), the familiar Erlang loss function. As can be seen, the curves verify the convexity of $G(x)$.

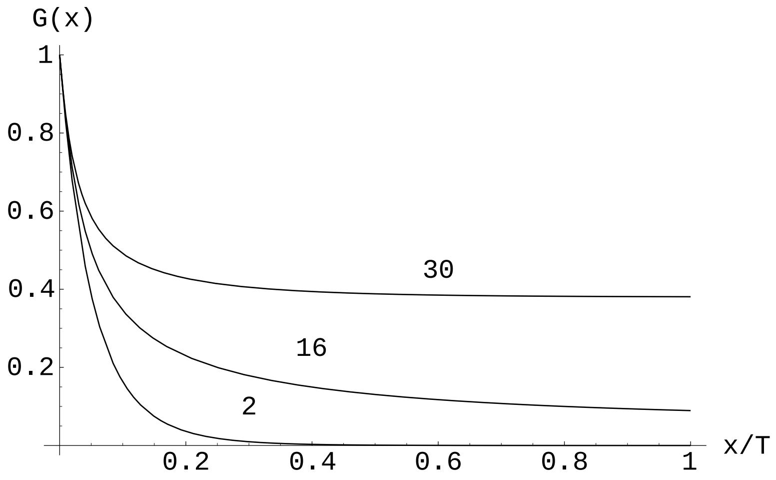

Figure 1: The probability $G(x)$ for the Erlang model with $c = 20$. The curves are labeled with values of ρ.

In the remainder of this section, we first approximate the probability of success for the Erlang loss model. We then calculate probabilities of success for the constant model when $n = 1$, and when n is arbitrary, $\tau = nT$ and T-spacing is used. The last two results lead to interesting conjectures.

4.1 Erlang loss model

Applications with large c are made difficult by the problem of finding the c roots of $R(s)$ in (6). The authors' version of MAPLE© was limited to $c < 27$. However, David Applegate has shown that Müller's method (see e.g. Conte and DeBoor [2]) applied to $R(s)$ gives the roots with c as high as 100.

For very rough calculations, the simple bound

$$G(x) \leq \frac{\rho + ce^{-(\rho+c)x/T}}{\rho + c} \tag{13}$$

can be useful as a conservative approximation. With $G_{c-1}(x)$ defined to be the conditional probability of state $c-1$ at time x given that the state at time 0 was c, the inequality $G_{c-1}(x) + G(x) \leq 1$ applied to the transition equation at state c,

$$\frac{dG(x)}{dx} = -\frac{c}{T}G(x) + aG_{c-1}(x),$$

can supply an easy proof of (13). As long as x is so small that the number of busy trunks is still highly likely to be $c-1$ or c, (13) gives a reasonably accurate approximation to $G(x)$. But eventually the bound becomes asymptotic to $\rho/(\rho + c)$ instead of to the true loss probability p_c from (1). Similar approximations, with k exponential terms, could be obtained by working with the transition equations for states $c, c-1, \ldots, c-k+1$. For further approximations, see Kosten [6].

By exploiting the special properties of the roots of $R(s)$, it might be possible to extend the calculations of $G(x)$ well beyond $c = 100$. For very large c, the asymptotic techniques of Mitra and Weiss [8] and Knessl [5] can also be recommended.

One of the asymptotic approximations introduced in [8] is as follows. Except when x is large, the most likely paths of the Erlang loss model from state c to state c in time x involve only a few transitions. Then the intermediate states k are all near c. At these states the transition rate $P(k \to k-1) = k/T$ may be approximated by c/T. With that approximation, and with states relabeled by the number $j = c - k$ of idle servers, the Erlang loss model is transformed into an $M/M/1$ queue with a buffer of size $c-1$. Arrivals (of idle servers) to the queue represent departures in the Erlang loss model and departures (ends of idleness) from the queue represent customer arrivals in the Erlang loss model. The transition rates for the queue are $P(j \to j+1) = c/T$ and $P(j \to j-1) = a$, except that $0 \to -1$ and $c \to c+1$ have probability 0. $G(x)$ becomes the probability that the queue is empty at time x, given that it was empty at time 0. With c large enough so that the queue's buffer is unlikely to be nearly full during time x, a further reasonable approximation replaces the finite buffer by one of infinite capacity. The probability that an $M/M/1$ queue in state 0 returns to state 0 in a time x is a standard result (see Riordan [9, p. 45, eq. (8)]). In our notation (Riordan's use of ρ and a is different), it is

$$\begin{aligned} G(x) &= e^{-(c+\rho)x/T}[I_0(2\sqrt{\rho}x/T) + \sqrt{\rho/c}I_1(2\sqrt{\rho}x/T) \\ &\quad + (1-c/\rho)\sum_{k=2}^{\infty}(\rho/c)^{k/2}I_k(2\sqrt{\rho}x/T)] \end{aligned} \tag{14}$$

where $I_k(z)$ is the Bessel function of imaginary argument $(-i)^k J_k(iz)$. The approximate $G(x)$ in (14) is often quite accurate. Thus, with $\rho = c = 20$, (14) is accurate to 1% for $0 \le x/T \le .08$, i.e., for $G(x) \ge .43$. Since the approximations used to derive (14) all increased the probabilities of transitions towards states with more servers idle, (14) is probably a lower bound on $G(x)$.

4.2 Constant busy periods, $n = 1$

In the exponential model, the probability of success increased with $\tau = X_n$ until, for large τ, the probability in (2) was obtained. That is no longer true with constant-length busy periods, as is made clear below just from the case $n = 1$.

Let ϱ be the residual lifetime of the call at time 0; ϱ is uniformly distributed on $(0, T)$. Before time τ, when the single retry is made, the called line can have some number k $(0 \le k < \tau/T)$ of other calls. The retry succeeds, for a given k, if and only if these k calls arrived and were served before time τ. Since the service time of k calls is kT, for a given value of $\varrho = r$, $0 < r < \min\{T, \tau - kT\}$, these k calls must arrive during the interval $\tau - r - kT$, an event having the Poisson probability distribution with mean $a(\tau - r - kT)$. The probability that the retry succeeds is then $\sum_{0 \le k \le \lfloor \tau/T \rfloor} P_k$, where

$$P_k = \int_0^{\min\{T, \tau-kT\}} \frac{[a(\tau - r - kT)]^k}{k!T} e^{-a(\tau-r-kT)} dr.$$

Integrations by parts yield

$$\rho P_k = H_k(a \max\{\tau - (k+1)T, 0\}) - H_k(a(\tau - kT)), \tag{15}$$

where

$$H_k(t) = e^{-t}[1 + t + t^2/2! + \cdots + t^k/k!]. \tag{16}$$

The terms P_k depend on τ in a way that changes its analytic form at $\tau = (k+1)T$. With $k = 0$ for example,

$$\rho P_0 = \begin{cases} 1 - e^{-a\tau} & \tau \le T, \\ e^{-a(\tau - T)} - e^{-a\tau} & \tau \ge T \end{cases} \tag{17}$$

Moreover, the number of terms P_k in the success probability depends on τ. As a result, the success probability depends on τ in a complicated way (see Figure 2 with $\rho = 2$). At $\tau = T$ the term P_0, which is always present, has a maximum exceeding $1/(1+\rho)$ by as much as 30% (depending on ρ). Then, with $n = 1$, the choice $\tau = T$ always improves on (2). Curves like Figure 2 for other values of ρ lead to the following

Conjecture 1. *In the model with constant busy periods and $n = 1$, a best choice of τ is $\tau = T$.*

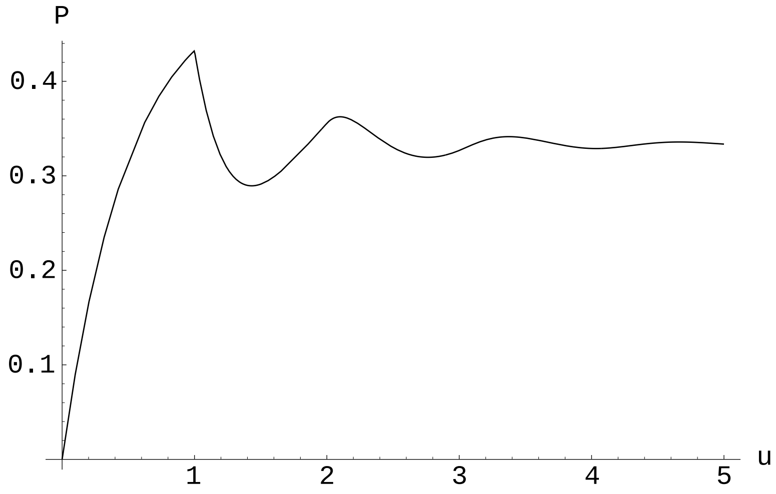

Figure 2: Probability P of success vs. $u = \tau/T$ for one retry and $\rho = 2$.

4.3 Constant busy periods, T-spacing

The T-spacing policy, with $X_k = kT$, is allowed if τ can be as large as nT. Calculations below show that this policy is better than either T/n-spacing or ∞-spacing; it may be optimal for $\tau = nT$, but that is not proved. Since calls all last for time T, the T-spacing policy fails only when, for $k = 1, \ldots, n$, the retry at X_k is blocked by a call that arrived in $((k-1)T, kT)$. The conditions that make T-spacing fail will be given in terms of the residual lifetime ϱ of the call at time 0 and the idle time y_k between the end of the $(k-1)$-st call and the start of the k-th call after time 0. When the policy fails, the retry at time kT was blocked by a call that started at time $Y_k = \varrho + y_1 + \cdots + y_k + (k-1)T$ and ended at time $Y_k + T$, where $Y_k < X_k = kT < Y_k + T$. All n retries fail if

$$\varrho + y_1 + y_2 + \cdots + y_k < T \tag{18}$$

for $k = 1, \ldots, n$. Since r and the y_i are all positive, all n inequalities (18) hold if (18) just holds at $k = n$.

In (18), ϱ is uniformly distributed on $(0, T)$ and y_i is exponential with mean $1/a$. For a given $\varrho = r$, and $k = n$, the $y_1, \ldots, y_n$ satisfy (18) with probability

$$1 - \sum_{i=0}^{n-1} \frac{[a(T-r)]^i}{i!} e^{a(T-r)}.$$

The policy's failure probability is obtained by averaging over r. The probability of success becomes

$$P(n, \tau) = \frac{n}{\rho} - \sum_{0}^{n-1} \frac{H_i(\rho)}{\rho}$$

with $H_i(t)$ as in (16). Terms of the sum can be combined to give the simpler result

$$P(n,\tau) = \frac{n}{\rho} - e^{-\rho} \sum_{i=0}^{n-1} (n-i) \frac{(\rho)^{i-1}}{i!}. \tag{19}$$

Table 1: Probabilities of success for three x-spacing policies

n	ρ	$x = T/n$	$x = T$	$x = \infty$
1	0.1	0.951626	0.951626	0.909091
1	0.3	0.863939	0.863939	0.769231
1	1	0.632121	0.632121	0.5
1	3	0.316738	0.316738	0.25
1	10	0.0999955	0.0999955	0.0909091
2	0.1	0.975412	0.998414	0.991736
2	0.3	0.928613	0.98706	0.946746
2	1	0.786939	0.896362	0.75
2	3	0.517913	0.583688	0.4375
2	10	0.198652	0.199946	0.173554
4	0.1	0.987604	0.999999	0.999932
4	0.3	0.96342	0.999945	0.997164
4	1	0.884797	0.995651	0.9375
4	3	0.703511	0.893548	0.683594
4	10	0.367166	0.398635	0.316987
7	0.1	0.992891	1	1
7	0.3	0.978874	1	0.999965
7	1	0.931855	0.999989	0.999023
7	3	0.813309	0.994269	0.866516
7	10	0.532244	0.675987	0.486842
10	0.1	0.995017	1	1
10	0.3	0.985149	1	1
10	1	0.951626	1	0.999023
10	3	0.863939	0.999872	0.943686
10	10	0.632121	0.87489	0.614457

Table 1 compares probabilities of success (12) and (19) against (2); the first two policies use even-spacing parameters $x = T/n, \tau = T$ and $x = T, \tau = nT$. For fixed n, the table shows that T-spacing is better than either T/n-spacing or ∞-spacing, especially when n and ρ are large. The results suggest the following

Conjecture 2. *In the model with constant busy periods and $\tau = nT$, the T-spacing policy is optimal.*

It is also interesting to see that Table 1 shows neither $x = T/n$ nor $x = \infty$ to be always better than the other.

5 Uneven-spacing policies with a fixed number of retries

This section shows that a policy using uneven spacing is appropriate when a is very small, as may be the case for a local call to a number without much traffic. Competition with other dialers is then not an important problem. If the dialer is willing to wait a time τ, a single retry at that time will come close to maximizing the probability of success. However, other trials before time τ might place the call with a shorter wait.

We also study a random policy that significantly simplifies calculation of probability of success for the Erlang loss model

5.1 No new arrivals

Suppose $a = 0$ and the dialer uses a policy of n retries, the last at time $X_n = \tau$. Any such strategy succeeds with probability $1 - G(\tau)$. Since $a = 0$, only a call in progress at time 0 can cause blocking. Then $G(x)$ becomes the probability that the call in progress has a residual lifetime x or more. With that interpretation, what follows applies even to the constant model; (3) is not used in this subsection.

In cases when the called number becomes free before time τ, the dialer now wants to succeed in the shortest mean time. The conditional probability of success at trial k, given a hang-up before time τ, is

$$G(X_k|\tau) = \frac{G(X_{k-1}) - G(X_k)}{1 - G(\tau)} \tag{20}$$

and the policy must minimize the conditional mean

$$\sum_{k=1}^{n} X_k G(X_k|\tau).$$

Minimizing conditions are obtained by setting derivatives with respect to X_k equal to zero. The result is a recurrence

$$[X_{k+1} - X_k]G'(X_k) = G(X_k) - G(X_{k-1}),\ k = 1, \ldots, n-1. \tag{21}$$

For the exponential model and $a = 0$, $G(x) = e^{-x/T}$ and $x_k = X_k - X_{k-1}$. Then (21) becomes

$$\frac{x_{k+1}}{T} = e^{x_k/T} - 1. \tag{22}$$

For given τ and n, the policy with smallest mean is found from (22). Starting with any trial value for x_1, (22) determines $x_2, \ldots, x_n$. The initial x_1 must be adjusted to make $x_n = \tau$. In a typical example, a dialer making $n = 4$ retries in time $\tau = 3T$ should dial at times $.456T$, $1.033T$, $1.815T$, and $3T$ to succeed in mean time $1.204T$. Although the best policy requires most of the retries to be made early, it doesn't improve much on τ/n spacing unless τ/T is large. In the example with $\tau/T = 3$ and $n = 4$, uniform spacing gives success in mean time $1.264T$. With larger values of n, any policy with reasonably closely distributed retry times will succeed almost immediately after hangup. The conditional mean time to hangup when $\tau = 3T$ is $.8428T$; the policy with $n = 4$ could be improved by increasing n.

With $a = 0$, models with more than one trunk have less interest. But one can adapt the above discussion to the Erlang model by noting that the first of c blocking calls to hang up has the same residual life as a single exponential call of mean duration T/c.

In the model with constant call durations, the residual-life distribution is

$$G(x) = 1 - x/T, \ \ 0 \le x \le T.$$

With $\tau \le T$ (the only reasonable condition), (21) now leads to the policy $X_k = k\tau/n$ instead of (22); the conditional mean wait is $(1 + 1/n)\tau/2$.

5.2 Random policy for the Erlang loss model

Instead of waiting a fixed interval between redials, a dialer might pick intervals $x_1, x_2, \ldots$ as i.i.d. choices of a random variable x. Indeed, random dialing may be better than perfectly regular dialing as a model of human behavior. Each retrial will now fail with probability $g = E(G(x))$. Since the x_k in (3) are independent, the random policy will succeed in n or fewer trials with probability $P(n) = 1 - g^n$. If $E(x) = y$, then

$$g = E(G(x)) \ge G(E(x)) = G(y) \tag{23}$$

follows from the convexity of $G(x)$ in (7). For a fixed mean y, no random policy does better than the y-spacing policy. A random dialer continuing until the call is placed uses a mean number $E(N) = 1/(1-g)$ of retries; they require a mean time $E(W) = y/(1-g)$. Suppose x has the exponential distribution, so that

$$g = g(y) = \int e^{-x/y} G(x) dx/y = L(1/y)/y \tag{24}$$

where $L(s)$ is the Laplace transform of $G(x)$. $L(s)$ was actually found as a preliminary step in the derivation of (7). Using that result, we now find immediately

$$g(y) = \frac{R(T/y - 1)}{R(T/y)} \tag{25}$$

with $R(s)$ the polynomial (6). Exponential spacing is then computationally simpler than constant spacing because it does not require solving for the zeros of (6). In fact, as (23) shows, $g(y)$ may serve as another bound on the more complicated $G(y)$.

To compare random with τ/n-spacing, consider the special case $c = 1$, i.e., the exponential model. Figure 3 compares the even-spacing success probability in (11) against the random-spacing success probability $1 - g^n(\tau/n)$, where $g(.)$ is computed from (24) after substituting (4). The integral gives

$$g(\tau/n) = \frac{1 + \rho\tau/(nT)}{1 + (1 + \rho)\tau/(nT)}$$

As a partial check of the results, it is easily verified that, as $n \to \infty$, both (11) and $1 - g^n(\tau/n)$ tend to $1 - e^{-\tau/T}$, the probability the blocking call ends in time τ.

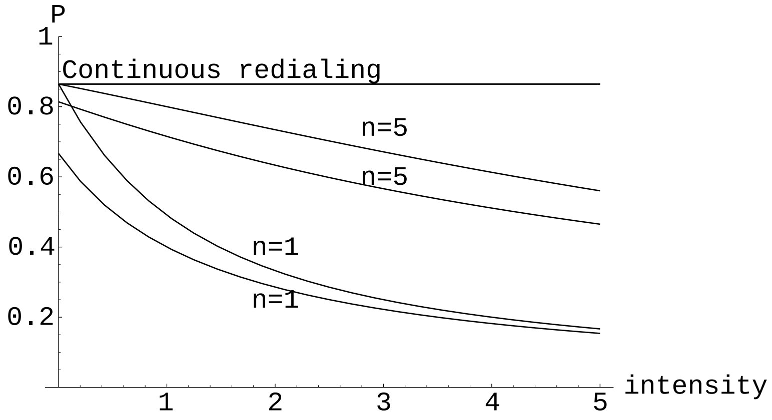

Figure 3: The success probability P vs. the traffic intensity ρ for the even and random (exponential) spacing policies with $\tau/T = 2$; the upper curve of each pair corresponds to the even spacing policy. In the limit $n \to \infty$, the two policies give the result labeled "continuous redialing" in the figure.

6 Certain success

Fixed numbers of retrials may be unacceptable to some dialers because failure may occur on all n trials. In this section we study policies requiring an automatic dialer to redial until it eventually succeeds.

6.1 Exponential and Erlang loss models

Assume that retries are made at times $x,\ 2x, \ldots$ until a successful trial is made. The number of retries actually used is a random variable ν. Again, each redial has probability $G(x)$ of failing. The dialer succeeds at trial n with probability $[1 - G(x)]G^{n-1}(x)$ so the expected number of trials is

$$E(\nu) = \frac{1}{1 - G(x)}$$

The dialer's mean wait to success is $E(W) = xE(\nu)$. Now the choice of x involves a compromise. Small x is needed for a short mean wait but large x is needed for a small mean number of redials (see Table 2).

The following criterion determines an interesting special value of x for the exponential model. Suppose that, when trials at $X_1, ..., X_{k-1}$ have failed, the next trial must maximize the probability of being the first call after the hang-up of the one in progress at time X_{k-1}. The choice of $x = X_k - X_{k-1}$ must maximize

$$\int_0^x e^{-t/T - a(x-t)} \frac{dt}{T} = \frac{e^{-x/T} - e^{-ax}}{\rho - 1}. \tag{26}$$

Table 2: Choosing x for the exponential model

ρ	x/T	$E(\nu)$	E(W)/T
0.2	0.2	5.62398	1.1248
0.2	0.5	2.65964	1.32982
0.2	1	1.71722	1.71722
0.2	2	1.31972	2.63945
0.2	10	1.20001	12.0001
1	0.2	6.06649	1.2133
1	0.5	3.16395	1.58198
1	1	2.31304	2.31304
1	2	2.03731	4.07463
1	10	2	20
5	0.2	8.58608	1.71722
5	0.5	6.31437	3.15719
5	1	6.01491	6.01491
5	2	6.00004	12.0001
5	10	6	60
25	0.2	26.1442	5.22885
25	0.5	26.0001	13
25	1	26	26
25	2	26	52
25	10	26	260

The maximum lies at $x/T = \ln(\rho)/(\rho - 1)$. Table 3 shows how this policy performs as a function of ρ.

Again, for the Erlang loss model, the dialer chooses x to balance $E(\nu)$ against $E(W)$. If x is large, then $E(\nu)$ is close to $1/(1 - p_c)$, with p_c the loss probability (see (1) with $k = c$), but $E(W)$ is large. If x is small, then $E(W)$ is near T/c, the mean wait for the first of c calls to end, but $E(\nu)$ is large. The dialer might now choose x to maximize the probability that in time x (i) one of the c calls ends and (ii) no new calls arrive afterward. Then a function like (26) (with T replaced by T/c) would be maximized; the choice would be

$$x = (T/c)\ln(\rho/c)/(\rho/c - 1), \tag{27}$$

and would be convenient because neither $G(x)$ nor the roots s_j are needed in (27).

6.2 No new arrivals

As before, the dialer might want a policy that is certain to succeed eventually. Suppose each retry is assumed to have a cost of b time units so that the cost to the dialer of

Table 3: Certain success policy with a special x

ρ	x/T	$E(\nu)$	$E(W)/T$
0.05	3.1534	1.08975	3.43642
0.1	2.55843	1.17015	2.99374
0.2	2.0118	1.31787	2.6513
0.5	1.38629	1.71429	2.3765
0.75	1.15073	2.01958	2.32399
1	1	2.31304	2.31304
1.25	0.892574	2.59881	2.31963
1.5	0.81093	2.87915	2.33479
2	0.693147	3.42857	2.3765
5	0.402359	6.58937	2.6513
10	0.255843	11.7015	2.99374
25	0.13412	26.8204	3.59715

success at trial k is $X_k + bk$. The X_k are to be chosen to minimize the expected cost

$$C = \sum_{k=1}^{\infty}[X_k + bk][G(X_{k-1}) - G(X_k)]. \tag{28}$$

(Note that k now runs from 1 to ∞.) A minimizing condition like (21) is

$$(X_{k+1} - X_k + b)G'(X_k) = G(X_k) - G(X_{k-1}), \tag{29}$$

or

$$\frac{x_{k+1} + b}{T} = e^{x_k/T} - 1 \tag{30}$$

for an exponential busy period. If the retry at X_1 fails, the call still has the same residual life distribution it had at time 0 and so the minimizing policy must have $x_1 = x_2$. That condition and (30) determine that all x_i are equal. The minimum C becomes

$$C = T + x_1 + b,$$

which is a low cost, considering that T is the mean cost of waiting for the call to end and b is the cost of one retry.

For a constant busy period, the minimizing condition (29) becomes

$$X_{k+1} - 2X_k + X_{k-1} + b = 0$$

with the solution

$$X_k = kX_1 - k(k-1)b/2. \tag{31}$$

In (31), the X_k will eventually decrease and so the policy will have to choose X_1 and a finite value of n such that $X_1 < X_2 < \cdots < X_n = T$. For a given n, $X_n = T$ requires $X_1 = T/n + (n-1)b/2$, so

$$n(n-1) < 2T/b$$

is needed to make $X_{n-1} < X_n$. In this way, the solution reduces to trying about $\sqrt{2T/b}$ values of n to see which gives the smallest C in (28). Unlike (22), equation (31) now clusters the X_k near T instead of spreading them evenly.

6.3 Constant busy periods, first call

As in Section 6.1, suppose a dialer always chooses X_k so that, given the failures at $X_1, \ldots, X_{k-1}$, his next call has maximum probability both of succeeding and of arriving before anyone else places a new call. The dialer's best policy is a T-spacing policy. A proof of this fact can easily be given assuming that Conjecture 1 in Section 4.2 is true. The argument that follows shows only the main idea.

By Conjecture 1, $x_1 = T$. To use an induction argument, suppose $x_1 = x_2 = \cdots = x_k = T$ and the first k retrials all fail. As in Section 4.3, the residual lifetime ϱ and the idle times y_i between calls satisfy (18). The call that blocked the trial at X_k has residual lifetime

$$Y_k + T - kT = \varrho + y_1 + y_2 + \ldots + y_k > \varrho,$$

which is even longer than the original residual lifetime of the call at 0. Since the best policy for a single retry waiting for a residual lifetime ϱ to end took $x_1 = T$, the policy to wait for an even longer lifetime should not take $x_{k+1} < T$. But, no lifetime exceeds T; a policy has nothing to gain by taking $x_{k+1} > T$. That leaves only $x_{k+1} = T$.

Acknowledgment. We are pleased to thank David Applegate for helpful discussions and for his guidance in our use of MAPLE©. We are also grateful to Alan Weiss for his comments on asymptotic approximations.

References

[1] V. E. Benes *Mathematical Theory of Connecting Networks and Telephone Traffic.* Academic Press, New York, 1965.

[2] S. Conte and C. DeBoor. *Elementary Numerical Analysis: An Algorithmic Approach*, 3rd Edition. McGraw–Hill, New York, 1980

[3] *Consumer Reports.* Consumer Union of the US, 101 Truman Ave. Yonkers, NY 10213-1053. Nov. 1995, 714–717.

[4] J. Haantjes. *Wiskundige Opgaven*, 17 (1938).

[5] C. Knessl. On the transient behavior of the M/M/m/m loss model. *Stochastic Models*, 6(1990), 749–776.

[6] L. Kosten. Over de invloed van herhaalde oproepen in de theorie der blockingskanen. *Die Ingenieur*, 59, pp. 1-25. (*Electrotechnik*, 14, E123–E131), 1947.

[7] W. Ledermann and G. E. H. Reuter Spectral theory for the differential equations of simple birth and death processes. *Phil. Trans. Roy. Soc.*, London, A236 321 (1954).

[8] D. Mitra and A. Weiss. The transient behavior in Erlang's model for large trunk groups and various traffic conditions. *Proceedings, Twelfth International Teletraffic Congress*, Torino, June 1988.

[9] J. Riordan. *Stochastic Service Systems.* Wiley, New York, 1962.