THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 61, No. 9, November 1982
Printed in U.S.A.

Evaluation of Private Networks

By R. L. KAUFMAN and A. J. M. KESTER

(Manuscript received March 17, 1982)

This paper describes algorithms for predicting end-to-end performance measurements in private networks. Service characteristics such as end-to-end blocking and delay are calculated based on point-to-point traffic data and a network routing guide. These techniques have been incorporated in the Enhanced Network Administration System. The system is routinely used by AT&T Long Lines and operating company network administrators for private network design and evaluation.

I. INTRODUCTION

An important procedure in network design is performance prediction. In a private network environment, a traffic engineer must recommend a network design that satisfies the customer's performance requirements. A set of computer programs called the Enhanced Network Administration System (ENADS) has been designed for administering Enhanced Private Switched Communications Service (EPSCS) networks and Electronic Tandem Switching (ETS) networks. The ENADS Network Service Evaluator (NETEVAL) uses point-to-point traffic data and a network routing guide to provide detailed service characteristics of a network design. This paper summarizes the NETEVAL algorithms.

In addition to characterizing network service, NETEVAL is used, in conjunction with other ENADS modules,[1] to ensure that the service characteristics chosen by the customer are achieved. The Network Synthesis (NETSYN) module designs a network that is close to the service specified by the customer. The service evaluator is used to ensure that the final network recommendation meets the customer's service requirements.

Traditionally, performance estimates of a large network in an alternate routing environment have been confined to a segment of a network, such as the intermachine trunk portion.[2] NETEVAL contains

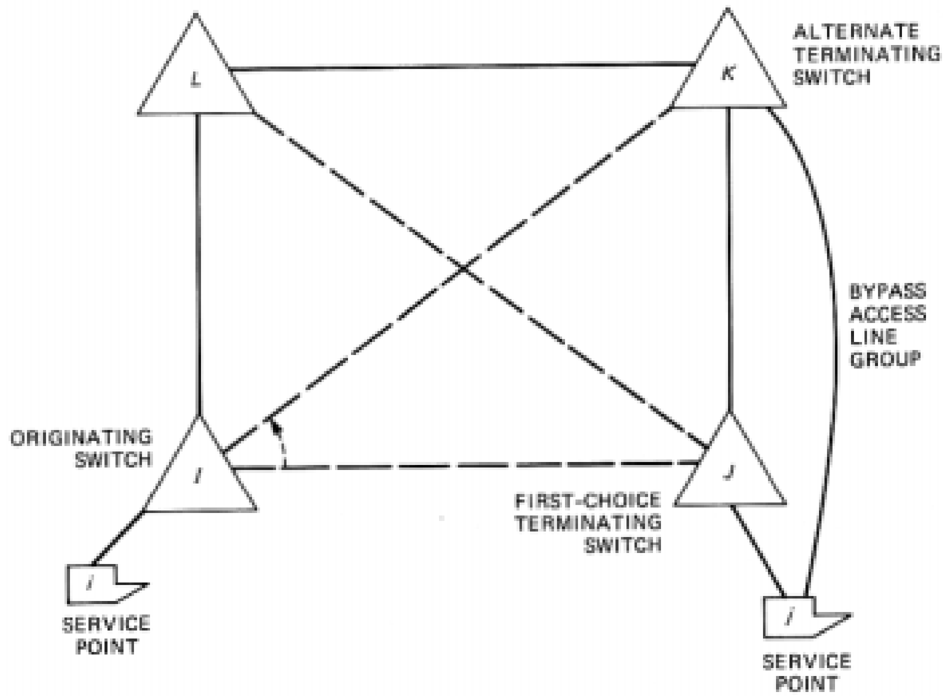

Fig. 1—Private network configuration.

decomposition and aggregation procedures that permit evaluation of complete end-to-end performance of a network as opposed to just a network segment. In addition, the ENADS evaluator generalizes the Katz algorithm to include private network features such as queues, and controlled access to network facilities via Facility Restriction Levels (FRLS). This paper describes the decomposition and aggregation procedures used in the evaluator, as well as the additional tools required to analyze some of the network features germane to private networks.

II. PRIVATE NETWORK CHARACTERISTICS

A private network (see Fig. 1) interconnects customer locations (on-net service points) and other locations (off-net service points). An on-net service point is generally a PBX/Centrex or a key set connected directly to an access line group. Off-net points are served by off-net facilities (Foreign Exchange, WATS, or local off-net access lines). In this paper both access lines and off-net facilities are called end-links. One end of each end-link is associated with a service point and the other end is connected (or homed) to a switch. A network of intermachine trunk groups connects the switches together, forming the intermachine trunk-group portion of the network. Switches permit concentration of point-to-point traffic.

Traffic generally originates at a service point, seizes an access line, and arrives at a switch. (It is possible in an ETS network for a service point to reside at the same location as the switch, so that an access line is not required.) NETEVAL assumes that the switch examines only the final destination of the traffic and its FRL and then associates an ordered route list of links at that switch. The route list is scanned until an idle circuit is found. If no such circuit exists, the call is blocked (in the absence of queuing). If a circuit is seized, the traffic arrives at the next switch, and the routing procedure continues until the call is either completed or blocked.

Other private network characteristics of interest for evaluation are queuing and FRLS. Queuing permits a call to wait at a specified trunk group for an idle circuit. The call is queued on a trunk group after a search through an ordered route list fails to find an idle circuit. The queued call may still be sent to reorder if the caller's wait exceeds a time-out threshold. Also, the caller may abandon the queue. The queue discipline is first come, first served.

FRLS provide the capability to restrict or expand access to network-facility route lists. The FRL can be based on the calling station and/or an authorization code. This provides the customer the opportunity to offer different grades of service by restrictive routing to different groups of users in the network. The FRL and final destination of a call are used in deciding to which links in the network the call will have access.

III. EARLIER SERVICE EVALUATION ALGORITHMS

Statistical techniques are available for estimating service characteristics of networks. It is well known that if first-offered traffic is generated from a Poisson process and holding times are exponentially distributed, then network characteristics can be obtained by formulating an appropriate Markov chain model and solving the resultant system of birth-death equations. This technique, however, requires an exorbitant amount of storage and computer time and is not considered feasible for large-scale networks.

Another reasonably accurate statistical technique that requires knowledge of only the traffic mean and variance (or the variance-to-mean ratio, called peakedness) is the Katz algorithm,[2] which was originally designed for evaluating switch-to-switch blocking probabilities. This algorithm requires estimates of switch-to-switch traffic means and variances that are "effectively" offered to each trunk group in the network.

Traffic that is carried on a link may be blocked on subsequent links. The holding time for such traffic on seized circuits is substantially less than that for completed traffic on the network. Effective-offered traffic reflects the shorter holding times of subsequently blocked traffic. The effective-offered load is used to compute the link-blocking probabilities and traffic-overflow variance for each link. The algorithm is an iterative process that updates effective-offered loads and link parameters at the end of each iteration.

After these initial link parameters have been computed, the switch-to-switch loads are distributed through the network based on the network-alternate routing plan and the link parameters. As this traffic is routed through the network, the effective means and variances of the traffic offered to each link are accumulated to update the link parameters for the subsequent iteration. The blocked traffic is also accumulated to provide probabilities for that iteration. After all the switch-to-switch loads have been distributed throughout the network, the link parameters are updated and the entire procedure is repeated until convergence in the switch-to-switch loss probabilities is obtained.

The Katz algorithm provides switch-to-switch blocking probabilities. Unfortunately, the number of calculations per iteration during the load assignment process is a function of the number of switches in the network. Traffic associated with each switch pair must be offered separately to each link in its routing path. Thus, an N-switch network must distribute $N(N-1)$ switch-to-switch loads through the intermachine trunk network. Since private networks generally consist of several hundred service points, it is not computationally feasible to define each service point as a switch and use the Katz algorithm. In addition, the Katz algorithm does not analyze queuing.

IV. NETEVAL ALGORITHM

NETEVAL is an iterative algorithm that successively updates traffic loads and associated traffic characteristics until a convergence criterion is satisfied. Because of the size of the network, NETEVAL decomposes the network into an end-link portion and an extended-trunk portion (see Section 4.4) to compute the traffic characteristics. After convergence the network performance components are aggregated to obtain end-to-end network performance estimates.

Sections 4.1 and 4.2 contain the NETEVAL assumptions and a brief outline of the algorithm. Subsequent sections explain in more detail the decomposition procedure, calculation of network parameters, and the aggregation techniques.

4.1 Model assumptions

The model assumptions for NETEVAL are as follows:

(*i*) Traffic means and variances provide a sufficient description of the loads.

(*ii*) The holding time on a link has four components: actual message time, ringing time, link set-up times, and subsequent queue delays.

Under this assumption blocked traffic can contribute positive loads to the network.

(*iii*) A blocked customer will redial with a specified retrial probability and a variance that is a fraction of the overflow variance.

(*iv*) No queue time-outs or abandonments occur.

(*v*) The system is in statistical equilibrium.

(*vi*) Only the final destination and FRL of a call are used at a switch to determine routing through the network.

4.2 NETEVAL algorithm for service evaluation

The basic algorithm procedure is given below. Some of the terms used are explained more clearly in later subsections.

(*i*) Decompose the network based on routing into an end-link network portion and an extended trunk portion.

(*ii*) Associate traffic parcels with each network portion.

(*iii*) Initialize all parcel blockings. (Zero can be used if no other estimates are available.)

(*iv*) Compute effective-offered parcels to the end-link network portion.

(*v*) Calculate blocking probabilities for the end-link parcels.

(*vi*) Compute total switch-to-switch and switch-to-final-destination parcels offered to the extended trunk portion of the network.

(*vii*) Calculate switch-to-switch and switch-to-final-destination blocking probabilities.

(*viii*) If end-link, switch-to-switch, and switch-to-final-destination blocking probabilities change significantly, return to step *iv*. Otherwise, go to step *ix*.

(*ix*) Compute point-to-point characteristics.

4.3 Decomposition

As we mentioned in Section III, a typical private network is too large to model each service point as a switch and use the Katz algorithm. The network must be decomposed into segments and analyzed separately. To define such a decomposition we must first define a traffic parcel.

The aggregate of point-to-point loads with identical routing, when offered to a particular portion of the network, will be called a traffic parcel for that network segment. For example, if all point-to-point traffic originating on service points homed on switch I and destined for service points homed on switch J have identical route lists when offered to the inter-machine trunk-group portion of the network, then such traffic forms a switch I to switch J parcel in the inter-machine trunk-group subnetwork. Any technique using such a traffic aggregation assumes that the network characteristics of the traffic parcel sufficiently approximate those of the individual point-to-point loads.

One method of decomposition is to separate the network into an inter-machine trunk-group portion and an end-link portion. Switch-to-switch parcels, as defined above, are associated with the inter-machine trunk-group network segment. All point-to-point traffic in the end-link segment, originating from a point to a switch and offered to the same route list of end-links, forms an originating end-link parcel. Similarly, all point-to-point traffic offered in the end-link network from a switch to a service point, using the same route list, is aggregated into a terminating end-link parcel.

Since each parcel's point-to-point load components have identical routing in its associated network segment, blocking can be computed for each parcel using the Katz algorithm. For example, switch-to-switch blocking can be computed for switch-to-switch parcels in the inter-machine trunk-group network segment, and originating and terminating parcel blockings can be computed in the end-link segment.

The above decomposition facilitates aggregation of parcel characteristics to obtain end-to-end characteristics. For example, if points i and j are homed on switches I and J, respectively, then i-to-j blocking (b_{ij}) is

$$b_{ij} = 1 - (1 - B^{I}_{i,\text{org}})(1 - B^{I,J})(1 - B^{J}_{j,\text{term}}). \tag{1}$$

(See Appendix A for a summary of the notations used.)

The three factors in the product represent the probability of call completion for the originating parcel i, the switch I to switch J parcel, and the terminating parcel j, respectively. An implicit assumption in such a decomposition is that all i-to-j traffic must use end-links homed on switch J for call completion. Or equivalently, the last switch, called the terminating switch, that completed i-to-j traffic encounters must be switch J. Figure 1, however, displays a typical private network in which terminating switches are not unique.

It illustrates a two-level hierarchical trunk-structure, with on-net service points homed on lower level switches I and J, respectively. A bypass access line group is used to route traffic from switch K to point j. The two-level trunk hierarchy permits i-to-j traffic to reach point j using either switch J or K as a terminating switch.

Since i-to-j traffic is not required to use the home access line group serving j, the above equation for b_{ij} is inaccurate. The last two terms in the product of the equation do not provide the switch I to point j blocking. To handle routing patterns with nonunique terminating switches, it is necessary to introduce a switch-to-final-destination parcel. Such a parcel represents all point-to-point loads from service points homed on a switch that has identical routing from the switch to the destination service point. In Fig. 1, a switch I to point j parcel is required to compute i-to-j blocking. If B^{Ij} is the blocking probability for such a parcel on the network, then b_{ij} can be accurately estimated by

$$b_{ij} = 1 - (1 - B^{I}_{i,\text{org}})(1 - B^{Ij}). \tag{2}$$

Thus, the NETEVAL decomposition must classify traffic into four parcels for adequate performance estimation: originating end-link parcels, terminating end-link parcels, switch-to-switch parcels, and switch-to-final-destination parcels. Switch-to-final-destination parcels are used only for service points that do not have unique terminating switches. To compute characteristics for switch-to-switch and switch-to-final-destination parcels, a segment of the network must be defined that contains all links used by these parcels. This subnetwork consists of all inter-machine trunk groups and those end-links contained in the route lists of the switch-to-final-destination parcels. Such a collection of links is called an extended trunk network in this paper. With the extended trunk network and its associated switch-to-switch and switch-to-final-destination parcel means and variances, the appropriate parcel characteristics can be computed. Similarly, the end-link subnetwork and end-link parcels permit the calculation of end-link parcel characteristics. Section 4.6 describes how parcel characteristics can be aggregated to form the desired end-to-end characteristics.

4.4 Link analysis

Both the end-link and switch-to-switch and switch-to-final-destination analyses are based on offering effective-offered loads to a single trunk group and computing associated link characteristics that are a function of overflow parcel means and variances. Whenever alternate routing is involved in a network segment, the Katz algorithm is used on the appropriate section. (Even in a queuing environment this procedure is valid except that a new algorithm to compute link parameters is used.) FRL analysis is included in NETEVAL by stratifying parcel loads by FRL and associating different route lists with each FRL grouping.

Effective-offered loads are computed for each link, from which link parameters such as blocking and overflow variance are estimated. The effective-offered loads associated with a link must take into account interactions with the rest of the network. Traffic blocking both prior and subsequent to a parcel being offered to a link reduces the effective-offered traffic load to the link. However, network setup times, ring times, queue delays, and retrial attempts increase effective-offered link loads. Even though the end-link and extended trunk network segments are analyzed separately in NETEVAL, the effective-offered load equa-

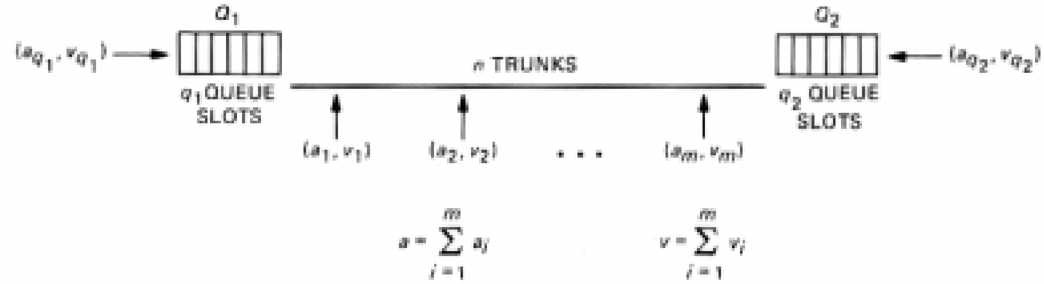

Fig. 2—Link configuration.

tions are designed to reflect the above network characteristics and thus account for the interactions between the network segments. These interactions are updated successively in the form of updated parcels in steps *iv* and *vi* as summarized in Section 4.2.

The following subsections describe the computation of link parameters both with and without queues.

4.4.1 Unqueued case

Figure 2 displays a general link configuration with link-offered parcels (a_i, v_i), $i = 1$ to m. a_i is the mean of the ith parcel and v_i is its variance. [If queues are present, the queue parcels are (a_{qi}, v_{qi}).] The equivalent random method can be used to compute overflow means and variances for aggregate traffic (a, v). To apportion the aggregate overflow moments to different parcels we use the Katz parcel-splitting approach. If b_i is the ith parcel blocking, then

$$b_i = b_o[1 + k(z_i - z_o)], \tag{3}$$

where,

b_o = aggregate link blocking,

z_o = peakedness of aggregate traffic,

$z_i = v_i/a_i$, and

k = modification factor.

Several heuristic formulas can be used for k. Katz[2] expresses k as a function of z_o and b_o. We have selected the heuristic formula from the Defense Communications Engineering Center

$$k = ce^{tn+sb_o}, \tag{4}$$

where $c = 2.249z_o^{-2.82}$, $t = 0.0528z_o^{-4.163}$, $s = 5.456z_o^{-2.025}$, and n is the number of trunks.

Once b_i is known, the parcel overflow and carried mean estimates are $a_i b_i$ and $a_i(1 - b_i)$, respectively. The overflow and carried variance

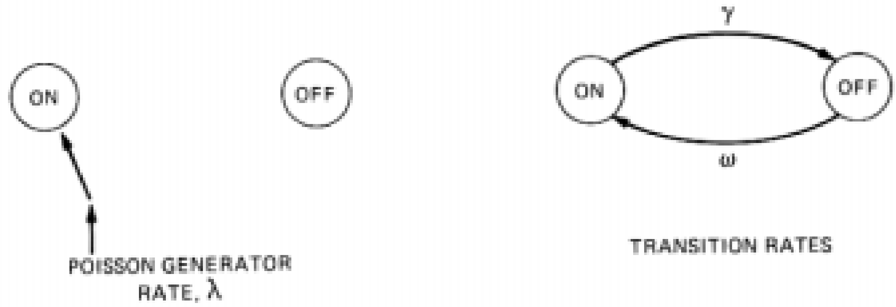

Fig. 3—The IPP process.

are approximated by $z_o^* a_i b_i$ and $v_i - z_o^* a_i b_i$, where z_o^* is the overall overflow peakedness. The overflow variance estimate results in equal peakedness to all parcels.

4.4.2 Queued case

To derive the overflow mean and variance for a parcel offered to a link in the presence of queues, another approach is needed. We approximate the behavior of a link containing queues with a two-dimensional Markov chain for each peaked parcel and a one-dimensional Markov chain for each smooth parcel. The parcel is modeled based on its mean and variance. If the parcel is peaked (peakedness greater than one), an Interrupted Poisson Process (IPP) is used. Smooth parcels (peakedness less than one), which can occur when traffic carried on previous groups is offered to a trunk group, are modeled as a Poisson process with parameters adjusted to better reflect the small peakedness.

4.4.2.1 Interrupted Poisson processes. An IPP, also called switched Poisson, is a Poisson process with rate λ, which alternately for some period of time shuts off all arrivals from the Poisson process, and lets the arrivals go through for another period of time. Both time periods are exponentially distributed, independent of each other and any previous time periods. IPP is used to model traffic with peakedness of one or greater. We say that the switch is ON when the arrivals go through, and otherwise the switch is OFF. The IPP process is shown in Fig. 3. A Poisson process is an IPP with the switch constantly in the ON position. Given (a_i, v_i), an IPP with an expected on time γ^{-1}, an expected off time ω^{-1}, and a Poisson rate λ (when the IPP is in the ON state) can be obtained from the relationships

$$a_i = \lambda \frac{\omega}{\omega + \lambda} \text{ and} \tag{5}$$

$$v_i - a_i + a_i^2 = a_i \lambda \frac{\omega + 1}{\omega + \lambda + 1}, \tag{6}$$

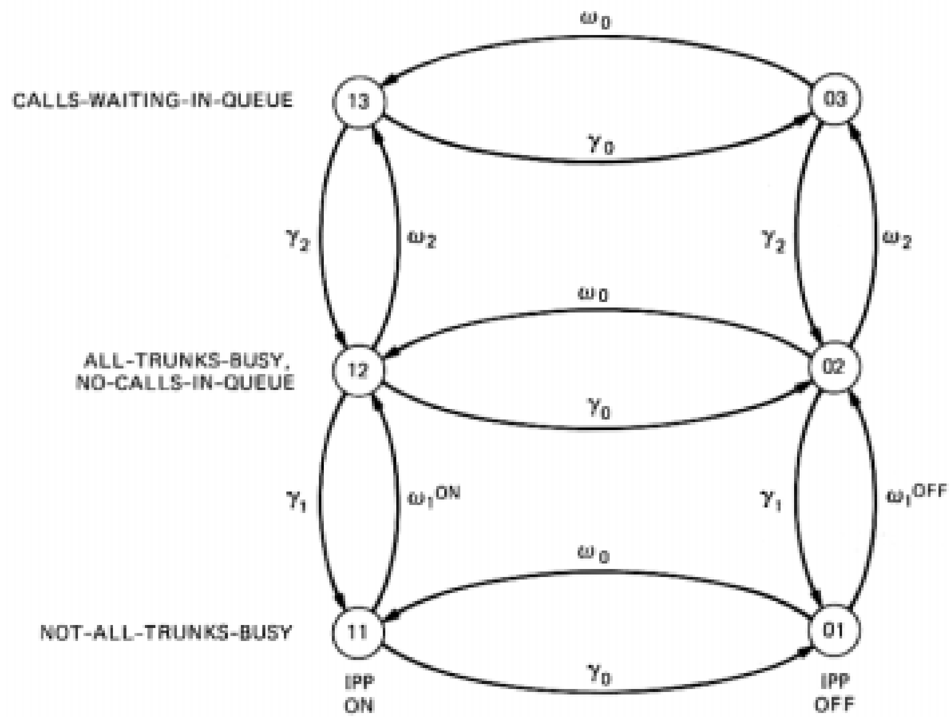

Fig. 4—Markov chain for peaked parcel input.

where λ is arbitrarily set to the larger of a_i and v_i. $v_i - a_i + a_i^2$ is the second factorial moment of the IPP.

This completely specifies the IPP point process. For further properties of the IPP, see Kuczura.[3]

4.4.2.2 Peaked parcel analysis. The IPP ON and OFF states form one dimension of the Markov chain used in the queued case. The other dimension consists of three superstates:

(*i*) Not-all-trunks-busy
(*ii*) All-trunks-busy, but no-calls-in-queue
(*iii*) Some-calls-waiting-in-queue.

Note that in case (*iii*), necessarily all trunks are busy, and that the system moves from state 1 only to state 2, from state 2 either to state 1 or to state 3, and from state 3 only to state 2. No transition between states 1 and 3 is possible.

Thus, the Markov chain space is (X, Y), $X = 0, 1$, $Y = 1, 2, 3$, where X is the IPP status of a parcel offered to a link and Y is one of the three superstates discussed above. Figure 4 displays the flows among the states.

Modeling the complex behavior of the trunk group with queues by compressing all states into superstates and assuming constant transition rates between them is a crucial assumption. The simplified state space is not Markovian any more, and the rates between the states are generally nonconstant. However, for first-moment calculations, such as the mean of the overflow stream, the replacement of the nonconstant transition rate by the average rate gives exact results. The calculation of the rates is given below. Some of the rates are themselves approximations for the average rate.

Since traffic is offered to the queues only when all trunks are busy in the group, we will consider the queue-offered traffic "conditional on all trunks busy," and use the conditional mean and variance of these streams when computing link characteristics. If (a_q, v_q) is the unconditional queue mean and variance, and p is the probability that all trunks are busy, then the mean and variance conditional on all-trunks-busy is given by

$$a_{cq} = a_q/p, \tag{7}$$

$$v_{cq} = \frac{v_q}{p} - \frac{a_q^2}{p}\left(\frac{1}{p} - 1\right). \tag{8}$$

These equations are based on the assumption that if the random variable X is the number of busy servers in an infinite trunk group receiving traffic from the queue stream, then $EX^i = pEX_c^i$, $i = 1, 2$.

4.4.2.3 Transition rates. To compute transition rates ω_1^{ON} and ω_1^{OFF}, we need to define ω_1 as the average frequency of transition from not-all-trunks-busy states to all-trunks-busy states. Since this rate does not depend on the presence of queues, we may calculate ω_1 on a trunk group without queues. Let b_T be the time congestion of a trunk group of size n, offered traffic with mean a and variance v. Further, let T_1 and T_2 be the expected sojourn times of the trunk-group system (without queues) in states not-all-trunks-busy and all-trunks-busy, respectively. In that case

$$b_T = \frac{T_2}{T_1 + T_2}. \tag{9}$$

Since $T_2 = n^{-1}$ (assuming unit average holding time) for known b_T, one may find T_1, and hence $\omega_1 = 1/T_1$, immediately. The time congestion b_T is equal to the blocking, b_i, that would be experienced by a Poisson parcel offered to the trunk group. This is given by the Katz parcel splitting formula

$$b_i = b_o[1 + k(z_i - z_o)], \tag{10}$$

where z_i is equal to 1.

Now that ω_1 is known we can solve for ω_1^{ON} and ω_1^{OFF}. We require that in the long run the number of transitions from states [(1, 1), (0, 1)] to [(1, 2), (0, 2)] in Fig. 4 is ω_1 per unit of time. Since movement among these states does not involve transition rates in or out of states [(1, 3), (0, 3)], it is more convenient to consider a smaller Markov chain

(X^*, Y^*), $X^* = 0, 1$, $Y^* = 1, 2$, which is a subset of the chain shown in Fig. 4. We can then express the relationship between ω_1, ω_1^{OFF}, and ω_1^{ON} as

$$\omega_1 = \frac{\rho_{01}}{\rho_{01} + \rho_{11}} \omega_1^{\text{OFF}} + \frac{\rho_{11}}{\rho_{01} + \rho_{11}} \omega_1^{\text{ON}}, \tag{11}$$

where ρ_{ij} is the steady-state probability distribution for the Markov chain (X^*, Y^*).

We still require an additional constraint to make ω_1^{OFF} and ω_1^{ON} unique. A reasonable additional assumption is that in the absence of queuing, the parcel blocking is the same as that obtained from the parcel splitting formula. Thus, we require

$$b_i = \frac{\rho_{12}}{\rho_{11} + \rho_{12}}, \tag{12}$$

where b_i is obtained from the Katz parcel splitting formula. The ρ_{ij} are themselves expressed in terms of ω_1^{OFF} and ω_1^{ON}. However, it can be shown that the equations above lead to an equation no worse than quadratic.

To compute the transition rate γ_1, assume the system is in an all-trunks-busy state, with no-calls-in-queue. γ_1 is the rate with which trunks become available. Since we express time in multiples of an average holding time, $\gamma_1 = n$ (the number of trunks). Further, under the assumption that the holding times are all independent, exponentially distributed variables, the transition rate γ_1 is a constant, independent of the time spent in the all-trunks-busy state, with no-calls-in-queue.

To compute ω_2 assume the system is in the state with all-trunks-busy and no-calls-in-queue. Then the frequency with which the state some-calls-waiting-in-queue is entered is the input rate into the queues. Since the queue-input rates a_{cq_1} and a_{cq_2} are given conditionally, it is clear that

$$\omega_2 = a_{cq_1} + a_{cq_2}, \tag{13}$$

where ω_2 is a constant, rather than the average rate when the queue input streams are both Poisson. If the input streams are peaked, they may be represented as interrupted Poisson processes, and a more precise, but more elaborate calculation for ω_2 may be carried out, taking account of the variances of the queue-input streams. This precision was not considered necessary.

Note that in all of the above, it is possible that one or both queues are non-existent. The mean a_{cq} should be set equal to 0 for non-existent queues.

γ_2, the rate from the state calls-in-queue to the state with all-trunks-busy and no-calls-in-queue, is the inverse of the mean length of time the system spends in the state with some-calls-in-queue. What we therefore need to calculate is the average duration from the time the first call enters one of the queues until the first subsequent time both queues are empty. We denote this mean duration by Q and call it "the queues busy period." One then has

$$\gamma_2 = Q^{-1}. \tag{14}$$

The calculation of Q is appreciably simpler for links with one queue than for links with two queues. In either case, one models the input stream(s) into the queue(s) as IPP(s), and then finds the simultaneous ergodic distribution of the number of waiting calls and the IPP input state. This ergodic distribution is a conditional distribution, given that all trunks are busy. The input streams are taken to be conditional on this event.

Lastly, ω_o and γ_o are the on and off rates for an IPP parcel offered to a trunk group. Thus, eqs. (5) and (6) can be used to compute ω_o and γ_o.

4.4.2.4 Computation of link characteristics. The overflow mean, overflow variance, and carried variance of a parcel (a_i, v_i) are computed by adding another dimension to the state space, which is the number of calls in an infinite trunk group that receives parcel overflow traffic when the system is in an all-trunks-busy state. Moment-generating functions can be used to compute the desired quantities. (Details are provided in Appendix B.) The overall link blocking is then the ratio of the sum of the parcel overflow means to the sum of the parcel offered loads.

Each of the parcels offered to the queues gives rise to two resultant parcels, namely the overflow parcel and the carried parcel. The queue overflow parcel consists of those calls finding all queue slots occupied, and the carried parcel contains all other calls, since we assume that there are no abandonments or time-outs.

As a result of assumed independence of the parcels offered to the group, the parcels offered to the two queues are independent during the time intervals that all trunks are busy. This means that waiting times and other variables associated with a two-queue trunk group may be estimated through simple Markov chain calculations.

Queue blocking and delay are computed conditional on all-trunks-busy. A Markov chain (j_1, j_2, i_1, i_2), where the j's are the number in each queue and the i's represent the on/off conditions of IPP queue streams, is used to compute these quantities. (Appendix C explains the algorithms used to compute the steady-state probabilities of the chain.) For example, if e_{j_1,j_2,i_1,i_2} is the ergodic distribution of the chain and (γ_1, ω_1) represents the on and off rates of the Q_1 arrival stream, then

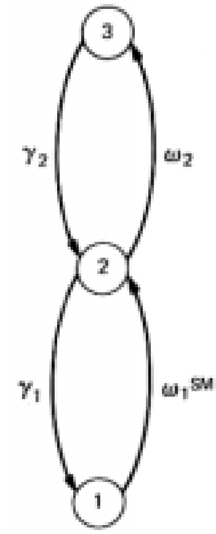

Fig. 5—Markov chain for smooth parcel input.

the Q_1 blocking probability is

$$b_{q_1} = \left(1 + \frac{\gamma_1}{\omega_1}\right) \sum_j (e_{q_1j10} + e_{q_1j11}). \qquad (15)$$

Since a customer can arrive at Q_1 only when the IPP is ON, eq. (15) is expressed only if the queue arrival stream turned on.

4.4.2.5 Smooth traffic parcel analysis. A smooth traffic parcel is offered to a link when v_i is less than a_i. A somewhat simplistic approach is used to compute the overflow stream characteristic for a smooth parcel. A one-dimensional Markov chain is formed, assuming the parcel is Poisson (see Fig. 5). All rates, except for ω_1^{SM} (used in place of ω_1^{ON} and ω_1^{OFF}), are the same as in the case of the peaked traffic analysis. ω_1^{SM} is computed in the same manner as ω_1, but with z_i set to the parcel peakedness.

Overflow quantities can be calculated using the same techniques applied to peaked parcels. The corresponding variance calculation, however, leads to too large a variance of the overflow, because the input into the trunk-group characterization switch is taken to be a Poisson stream with mean and variance (a_i, a_i), instead of the given smooth stream with mean and variance (a_i, v_i). To correct for this effect, we reduce the calculated variance by multiplying it by the peakedness (less than one), of the offered parcel. The estimated overflow variance is $z_i v_i^{o*}$, where v_i^{o*} is the overflow variance based on Poisson input. The mean overflow is given by

$$a_i \frac{\omega_1^{SM}(\omega_2 + \gamma_2)}{(\omega_1^{SM} + \gamma_1)(\omega_2 + \gamma_2) - \gamma_1\omega_2}. \qquad (16)$$

4.5 Convergence criteria

The convergence criteria for step *viii* of the NETEVAL algorithm (Section 4.2) assumes convergence has occurred if the absolute difference of each end-link, switch-to-switch, and switch-to-final-destination blocking probability in successive iterations is within a specified limit.

4.6 Point-to-point characteristics

Once the network has been decomposed into the extended trunk network and the end-link network, and the components have been analyzed, aggregation techniques are required to compute point-to-point characteristics. This is done by associating two or three traffic parcels, depending on the routing patterns, with each point-to-point pair. If i-to-j traffic is part of a switch-to-final-destination parcel when offered to the extended trunk network, then only two parcels are associated with the point pair: the originating parcel at point i offered to the end-link network segment, and the switch-to-final-destination parcel in which it is contained. If the i-to-j traffic is part of a switch-to-switch parcel in the extended network, then three traffic parcels are associated with the point pair: the originating parcel, the switch-to-switch parcel, and the terminating parcel in which i-to-j traffic is contained. Point-to-point characteristics are functions of the corresponding parcel characteristics.

If i-to-j traffic is part of a switch-to-switch parcel when traversing the extended trunk network, the i-to-j blocking b_{ij} is

$$b_{ij} = 1 - (1 - B^I_{i,\mathrm{org}})(1 - B^{IJ})(1 - B^J_{j,\mathrm{term}}). \tag{17}$$

If i-to-j traffic, however, is contained in a switch-to-final-destination parcel,

$$b_{ij} = 1 - (1 - B^I_{i,\mathrm{org}})(1 - B^{Ij}). \tag{18}$$

The i-to-j expected queue delay, W_{ij}, is a weighted average of the expected queue delay on trunks and end-links. When i-to-j traffic is contained in a switch-to-switch parcel,

$$W_{ij} = (1 - B^I_{i,\mathrm{org}})W^{IJ} + (1 - B^I_{i,\mathrm{org}})(1 - B^{IJ})W^J_j. \tag{19}$$

The trunk delay, W^{IJ}, and terminating link delay, W^J_j, are weighted by the probability that i-to-j traffic will be offered to a link on which it is queue-eligible. When i-to-j traffic is contained in a switch-to-final-destination parcel,

$$W_{ij} = (1 - B^I_{i,\mathrm{org}})W^{Ij}, \tag{20}$$

where W^{Ij} is the expected queue delay for traffic originating at switch I destined for point j. $1 - B^I_{i,\mathrm{org}}$ is the probability that the traffic seizes an end-link and arrives at switch I.

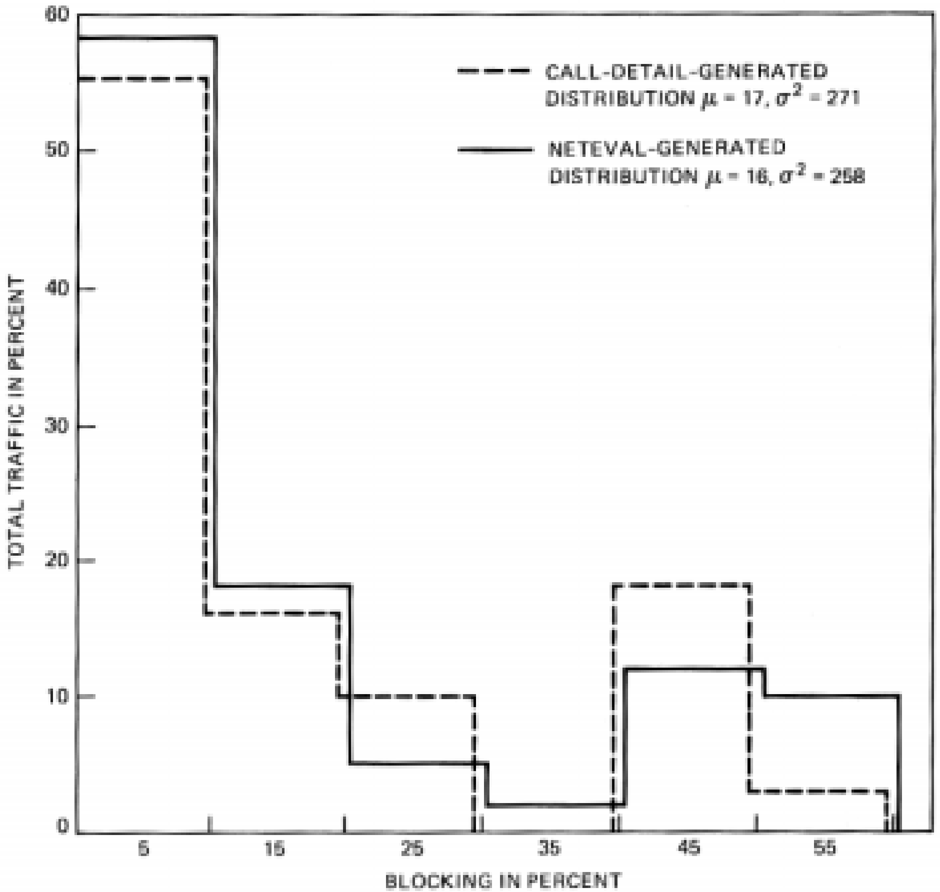

Fig. 6—End-to-end blocking distribution—Network A.

The i-to-j queue delay probabilities, PD_{ij}, are computed analogous to W_{ij}. When i-to-j traffic is contained in a switch-to-switch parcel when traversing the extended trunk network,

$$PD_{ij} = (1 - B^I_{i,\text{org}})PD^{IJ} + (1 - B^I_{i,\text{org}})(1 - B^{IJ} - PD^{IJ})PD^J_j, \quad (21)$$

where $(1 - B^I_{i,\text{org}})(1 - B^{IJ} - PD^{IJ})PD^J_j$, is the probability of traversing the trunk portion of the network without queue delay but incurring queue delay on the terminating end-link.

If i-to-j traffic is contained in a switch-to-final-destination parcel,

$$PD^{ij} = (1 - B^I_{i,\text{org}})PD^{Ij}. \quad (22)$$

Conditional expected queue delays are given by W_{ij}/PD_{ij}, for $PD_{ij} > 0$. Note that by creating an artificial access-line group with no blocking or delay, the above equations remain valid for ETS networks with traffic originating at a tandem switch.

V. APPLICATIONS

EPSCS and ETS networks generate call-detail records. These records contain information on individual calls in the network from which it is

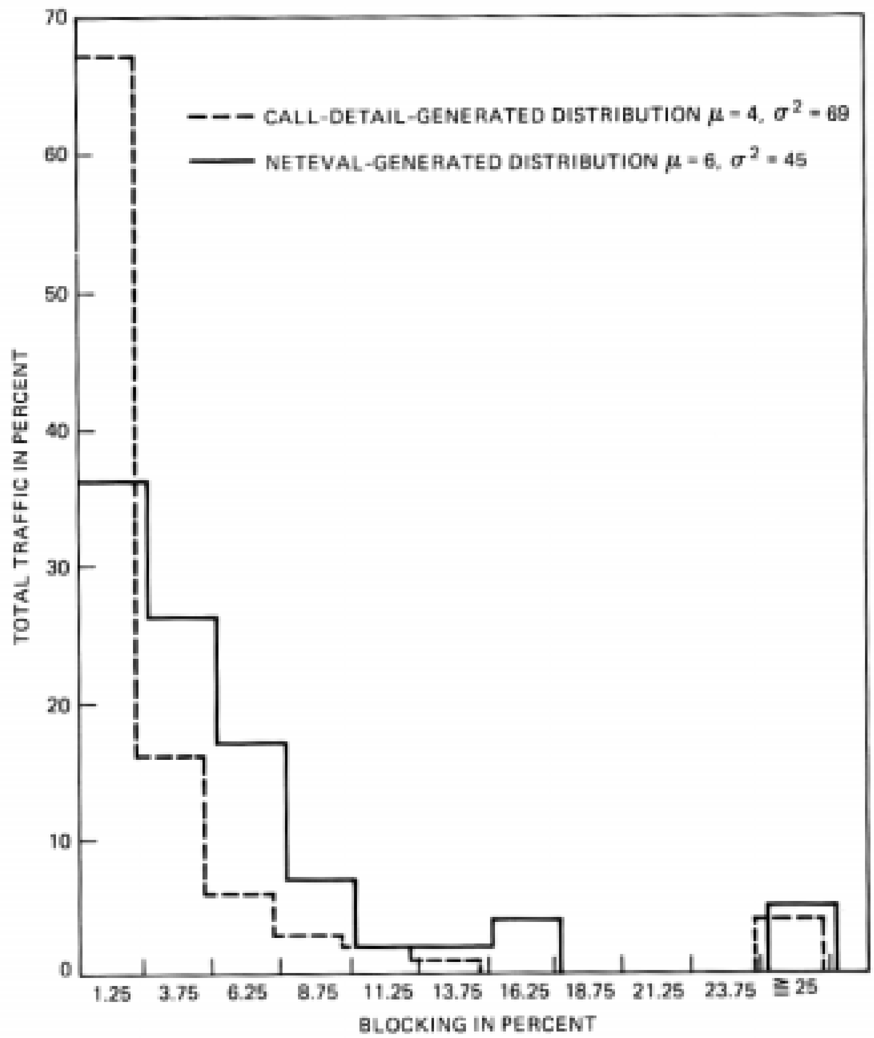

Fig. 7—End-to-end blocking distribution—Network B.

possible to estimate the end-to-end blocking distribution in existing networks. This information can be compared with an ENADS NETEVAL analysis of the same network. Figures 6 and 7 display both call detail and NETEVAL-derived blocking distributions for two existing networks.

The means and variances in Fig. 6 are very close. The variances in Fig. 7 are not as similar as their corresponding means. This is primarily because the midpoint of the last cell in the call-detail-derived distribution is 43 percent, compared with 30 percent in the NETEVAL distribution. Note, however, that the overall shape of the distributions within Figs. 6 and 7 is similar. This suggests that the modeling techniques in our evaluation algorithm provide a reasonable prediction of end-to-end performance.

VI. SUMMARY

Algorithms for computing end-to-end and link characteristics in private networks have been described in this paper. The algorithms generalize the Katz procedure to provide end-to-end characteristics for large-scale networks and also model private network features such as queuing and FRLs. These techniques have been incorporated into the ENADS Service Evaluator. NETEVAL performance predictions have compared well with actual measurements for several networks. The ENADS NETEVAL module is now routinely used with the companion NETSYN module by AT&T Long Lines and Operating Company private network administrators for network design and evaluation.

VII. ACKNOWLEDGMENTS

The authors acknowledge the preliminary evaluation work of J. K. Andreozzi and the end-to-end blocking distributions provided by W. E. Relyea.

REFERENCES

1. A. L. Kalro, "Traffic Network Design of New Private Switched Services," Telecommun. Networks Planning Symp., Paris, France, 1980.
2. S. S. Katz, "Statistical Performance Analysis of a Switched Communications Network," paper presented at the 5th ITC Congress in New York, 1967.
3. A. Kuczura, "The Interrupted Poisson Process as an Overflow Process," B.S.T.J., *52*, No. 3 (March 1973), pp. 437–48.

APPENDIX A

Notations

Much of the notation used in Section IV is summarized below for the reader's convenience.

Parcel Characteristics

Blocking

b_{ij} = Point i to point j blocking.

$B^I_{i,\text{org}}$ = Originating end-link blocking for service point i homed on switch I.

$B^J_{j,\text{term}}$ = Terminating end-link blocking for service point j homed on switch J.

B^{IJ} = Switch I to switch J blocking.

B^{Ij} = Switch I to final destination point j blocking.

Delays

W_{ij} = Point i to point j queue delay.

W^J_j = Terminating end-link queue delay for service point j homed on switch J.

W^{IJ} = Switch I to switch J queue delay.

W^{Ij} = Switch I to final destination point j queue delay.

PD_{ij} = Point i to point j delay probability.

PD_j^J = Terminating end-link delay probability for service point j homed on switch J.

PD^{IJ} = Switch I to switch J delay probability.

PD^{Ij} = Switch I to final destination point j delay probability.

Link Parameters

(a_i, v_i) = Mean and variance of ith parcel offered to trunk group.

z_i = Peakedness of ith parcel offered to trunk group.

(a_{q_i}, v_{q_i}) = Mean and variance of load offered to queue.

(a_{cq_i}, v_{cq_i}) = Mean and variance of load offered to queue conditioned on all trunks busy.

b_i = Blocking of ith parcel offered to trunk group.

(a, v) = Aggregate trunk-group load mean and variance.

z_o = Trunk-group peakedness of offered load.

z_o^* = Trunk-group overflow peakedness.

b_o = Trunk-group blocking.

Markov Transition Rates in Trunk-Group Queuing Model

γ_o = Interrupted Poisson Process off rate.

ω_o = Interrupted Poisson Process on rate.

ω_1^{ON} = Not-all-trunks-busy to all-trunks-busy, no-calls-in-queue flow rate, given Interrupted Poisson Process (IPP) on.

ω_2 = All-trunks-busy, no-calls-in-queue to calls-waiting-in-queue flow rate.

ω_1^{OFF} = Not-all-trunks-busy to all-trunks-busy, no-calls-in-queue flow rate, given IPP off.

γ_2 = Calls-waiting-in-queue to all-trunks-busy, no-calls-in-queue flow rate.

γ_1 = All-trunks-busy, no-calls-in-queue to not-all-trunks-busy flow rate.

APPENDIX B

Overflow Mean and Variance of Peaked Parcel

This appendix gives the details of the calculation of the overflow mean and variance of a parcel, modeled as an IPP, offered to a trunk group, which in turn is modeled by three states and the four transition rates among them. Let the variables λ, γ_0, γ_1, γ_2, ω_0, ω_1^{ON}, ω_1^{OFF} and ω_2 be as in Section 4.4.2.

The system to be treated, including the infinite trunk group, is a continuous-time Markov chain, with state space

$$N = \{(k, i, j): k = 0, 1, \cdots ; i = 0, 1; j = 1, 2, 3\},$$

where k is the number of calls in the infinite trunk group, i is the status (ON = 1, OFF = 0) of the IPP, and j is the state of the trunk group, (not-all-trunks-busy, all-trunks-busy, but no-calls-in-queue, some-calls-waiting-in queue). Let S_1 denote the status of the IPP and S_2 the status of the trunk group. Designating by p_{kij} the steady-state probability that the system is in state (k, i, j), we define the generating functions

$$\pi_{ij}(s) = \sum_{k=0}^{\infty} p_{kij} s^k, \qquad \text{for} \quad 0 \le s \le 1, \qquad i = 0, 1, \qquad j = 1, 2, 3.$$

The mean number of calls in the infinite trunk group is

$$a^* = \sum_{i,j} \sum_{k} k p_{kij} = \sum_{i,j} \frac{d}{ds} \pi_{ij}(1) = 1' \frac{d}{ds} \pi(1),$$

where $1'$ stands for the six element row vector of ones and

$$\pi(s) = \begin{vmatrix} \pi_{01}(s) \\ \pi_{02}(s) \\ \pi_{03}(s) \\ \pi_{11}(s) \\ \pi_{12}(s) \\ \pi_{13}(s) \end{vmatrix} .$$

Similarly, the second factorial moment is

$$EX(X-1) = \sum_{i,j} \sum_{k} k(k-1) p_{kij} = \sum_{i,j} \frac{d^2}{ds^2} \pi_{ij}(1) = 1' \frac{d^2}{ds^2} \pi(1),$$

from which one easily finds the variance. Writing out the steady-state equations for p_{kij} and summing appropriately, one finds that the $\pi_{ij}(s)$ satisfy a system of equations, which we will define in vector form. We define

$$A = \begin{vmatrix} -(\omega_0 + \omega_1^{\mathrm{OFF}}) & \gamma_1 & 0 & \gamma_0 & 0 & 0 \\ \omega_1^{\mathrm{OFF}} & (\omega_0 + \gamma_1 + \omega_2) & \gamma_2 & 0 & \gamma_0 & 0 \\ 0 & \omega_2 & -(\omega_0 + \gamma_2) & 0 & 0 & \gamma_0 \\ \omega_0 & 0 & 0 & -(\gamma_0 + \omega_1^{\mathrm{ON}}) & \gamma_1 & 0 \\ 0 & \omega_0 & 0 & \omega_1^{\mathrm{ON}} & -(\gamma_0 + \gamma_1 + \omega_2) & \gamma_2 \\ 0 & 0 & \omega_0 & 0 & \omega_2 & -(\gamma_0 + \gamma_2) \end{vmatrix},$$

and the diagonal matrix B by

$$B = \mathrm{diag}(0, 0, 0, 0, \lambda, \lambda),$$

where the elements of the diagonal are given inside the parentheses.

The vector $\pi(s)$ satisfies the equation

$$(s-1) \frac{d}{ds} \pi(s) = A\pi(s) + (s-1)B\pi(s). \tag{23}$$

Setting $s = 1$ in (23), one obtains $A\pi(1) = 0$, from which the marginal probabilities $\pi_{ij}(1) = P\{S_1 = i, S_2 = j\}$ are found. Taking derivatives w.r.t. s in (23) and setting $s = 1$ gives

$$(I - A)\frac{d}{ds}\pi(1) = B\pi(1).$$

Since $1'A = 0$, one gets

$$1'\frac{d}{ds}\pi(1) = 1'(I - A)\frac{d}{ds}\pi(1) = 1'B\pi(1), \tag{24}$$

and hence the mean α^* of the overflow may be found by solving for $A\pi(1) = 0$, with the side condition $1'\pi(1) = 1$. Differentiating eq. (23) twice, one gets

$$2\frac{d^2}{ds^2}\pi(s) + (s-1)\frac{d^3}{ds^3}\pi(s) = A\frac{d^2}{ds^2}\pi(s) + 2B\frac{d}{ds}\pi(s) + (s-1)B\frac{d^2}{ds^2}\pi(s).$$

Setting $s = 1$ and solving for $\frac{d^2}{ds^2}\pi(1)$ gives

$$\frac{d^2}{ds^2}\pi(1) = 2(2I - A)^{-1}B\frac{d}{ds}\pi(1).$$

Again, since $1'(2I - A) = 2\,1'$, one finds

$$1'\frac{d^2}{ds^2}\pi(1) = 1'B\frac{d}{ds}\pi(1),$$

or, substituting (24),

$$1'\frac{d^2}{ds^2}\pi(1) = 1'B(I - A)^{-1}B\pi(1).$$

Thus, by inverting the 6×6 matrix $I - A$, the variance of the overflow may be found.

The carried traffic mean may be obtained directly from the overflow mean via

$$\alpha_c = \alpha - \alpha^*,$$

where α_c is the carried mean, α the offered traffic mean, and α^* the overflow mean. It can be shown that α_c may alternatively be found via the procedure sketched above, if we change the B matrix to

$$B = \text{diag}(0, 0, 0, \lambda, 0, 0).$$

This change represents the fact that carried traffic leaves when the IPP switch is ON, and the trunk group characterization is in the not-all-trunks-busy state.

With this change in B, the variance of the carried traffic may be calculated in exactly the same way as the overflow variance.

APPENDIX C

Queue Analysis

This appendix gives some details of the queue analysis. Assume a trunk group with n trunks, q_1 queue slots on one side, q_2 on the other, and input streams into the queues given, conditionally on all-trunks-busy, by their IPP parameters $(\lambda_i,\ \omega_i,\ \gamma_i)$, $i = 1, 2$. The system is a continuous time Markov chain. This appendix gives a description and a fixed ordering of the states of the system in Section C.1 and algorithms for the steady-state probabilities in Section C.2.

C.1 Ordering of the states

The elements of the different vectors and matrices to be used in this appendix are characterized by a four-dimensional index, $(i,\ j,\ o1,\ o2)$, where each possible index is a state of the system. Index i is the number of calls waiting in the first queue, j the number of calls waiting in the second queue, and $o1$ and $o2$ are status bits, indicating whether the IPP representing the input stream is currently in the OFF state ($o = 0$) or in the ON state ($o = 1$). Status bit $o1$ refers to the first queue, $o2$ to the second. Index i runs between 0 and q_1, j between 0 and q_2. All bounds are inclusive.

A fixed order of enumeration is adhered to. There is a macro ordering, for the $(i,\ j)$ part of the state designations, and within that ordering, the status bits have a fixed order.

The gross order is given by

$$
\begin{array}{ccccc}
(0,0), & (1,0), & (2,0), & \cdots & (q_1,0),\\
(0,1), & (1,1), & \cdot & \cdots & (q_1,1),\\
(0,2), & \cdot & \cdot & \cdots & \cdot\\
\cdot & \cdot & \cdot & \cdots & \cdot\\
(0,q_2), & (1,q_2), & \cdot & \cdots & (q_1,q_2).
\end{array}
$$

Within these index elements there are four states indicating the status bits, which are ordered like 00, 01, 10, 11. Thus, the overall order is indicated by 0000, 0001, 0010, 0011, 1000, 1001, 1010, 1011, 2000, etc.

C.2 Steady-State Equations

The steady-state equations form a set of $4\,(q_1 + 1)\,(q_2 + 1)$ equations in as many unknowns, with the normalizing side conditions that the probabilities add to 1. In the ordering given above, the structure of these equations is that of a tri-diagonal blocked matrix equation, as follows. Let e_j be the vector of ergodic probabilities associated with the states having j calls in Q_2, $j = 0, 1, \cdots, q_2$, i.e.,

$$e_j = (e_{0j00},\ e_{0j01},\ e_{0j10},\ e_{0j11},\ e_{1j00},\ \cdots,\ e_{q_1j11}).$$

Then the e_j satisfies

$$e_0 A_e + e_1 C = 0, \tag{25a}$$

$$e_{j-1} B + e_j A_g + e_{j+1} C = 0, \qquad j = 1, \cdots, q_2 - 1 \tag{25b}$$

$$e_{q_2-1} B + e_{q_2} A_f = 0. \tag{25c}$$

Writing out these equations in matrix form will show the tri-diagonal block structure.

The matrices A_e, A_f, A_g, B, and C are all square of order $4(q_1 + 1)$; in addition, B and C are both diagonal matrices. Below we discuss each of these matrices, and exhibit their structures.

(*i*) C is associated with a call arriving into Q_2: the system moves from state (i, j, o_1, o_2) to state $(i, j + 1, o_1, o_2)$, with transition rate λ_2 if and only if $o_2 = 1$. Therefore, $C = \text{diag}(0, \lambda_2, 0, \lambda_2, \cdots, 0, \lambda_2)$.

(*ii*) B signifies the transitions caused by a call waiting in Q_2 being served. The transition is from (i, j, o_1, o_2) to $(i, j - 1, o_1, o_2)$, with rate n if $i = 0$, and with rate n_2 (where n_2 is the effective number of circuits available to calls from Q_2 when Q_1 and Q_2 are both non-empty) if $i \neq 0$. Specifically, $B = \text{diag}(n, n, n, n, n_2, n_2, \cdots, n_2)$.

(*iii*) A_e, A_f, and A_g are all associated with transitions for which the number of calls in Q_2 is constant, this constant being equal to 0 for A_e (empty), q_2 for A_f (full), and general for A_g. These matrices have a tri-diagonal block structure themselves, with blocks of size 4×4, with the super- and sub-diagonal blocks associated with calls entering Q_1 and leaving Q_1, respectively, and with the diagonal blocks modeling the transitions between the input status states $(o_1, o_2) = (0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$. A_e, A_f and A_g have slight differences owing to the fact that

(*a*) The diagonal element for any state contains the negative sum of the transition rates out of that state.

(*b*) The rate at which calls leave Q_1 depends on whether Q_2 is empty.

Next we show that we may take advantage of the block structure of the $4(q_1 + 1)(q_2 + 1)$ set of steady-state equations, and reduce it to a set of size $4(q_1 + 1)$ to be solved via brute-force matrix inversion methods. Consider equations (25). Since B is a nonsingular diagonal matrix, we may write

$$e_{q_2-1} = -e_{q_2} A_f B^{-1}$$

$$e_{q_2-2} = -e_{q_2}(CB^{-1} + A_f B^{-1} A_g B^{-1}).$$

In general, if $e_j = e_{q_2} D_j$ for $j = k + 1, k + 2, \cdots, q_2$, one obtains D_k from equation (25b)

$$D_k = -D_{k+1} A_g b^{-1} - D_{k+2} C B^{-1} \qquad (k = 0, 1, \cdots, q_2 - 1),$$

with $D_{q_2} = I$, the unit matrix, and $D_{q_2+1} = 0$.

Assume then, that we have expressed e_0 in terms of e_{q_2}. Now we use the top equation to solve for e_{q_2} and get

$$e_{q_2}(D_0A_e + D_1C) = 0. \tag{26}$$

The requirement that all probabilities sum to 1 can likewise be expressed in terms of e_{q_2} only:

$$\Sigma e_j 1 = \Sigma e_{q_2} D_j 1 = 1.$$

Thus, we sum the rowsums of the D_j, obtaining

$$x = \Sigma D_j 1.$$

We now replace one column of $D_0A_e + D_1C$ by the vector x, and the corresponding entry in the 0 vector in (26) by 1, and solve this modified set of equations. We obtain e_{q_2}, and from it may find all ergodic probabilities by postmultiplying e_{q_2} by the D_j's.

If the number of queue slots for $Q1$ is larger than for $Q2$ ($q_1 > q_2$), we may choose to relabel $Q1$ and $Q2$, and reduce the problem size by performing the recursive procedure described above over the larger of q_1 and q_2. Thus, we may reduce the size of the matrix to be inverted from $4(q_1 + 1)(q_2 + 1)$ to $4[1 + \min(q_1, q_2)]$.