PERFORMANCE EVALUATION OF VARIABLE-BIT-RATE VOICE IN PACKET-SWITCHED NETWORKS

V. R. Karanam, K. Sriram, and Duane O. Bowker

V. R. Karanam, K. Sriram, *and* ***Duane O. Bowker*** *are with AT&T Bell Laboratories in Holmdel, New Jersey. Mr. Karanam is a member of technical staff in the Network Performance Planning Department. He is involved in voice performance planning, with emphasis on evaluation of emerging network technologies. He received a B.E. in electronics and communication engineering from Osmania University, India, and M.S. and Ph.D. degrees in electrical engineering from Oregon State University. He joined AT&T in 1983. Mr. Sriram is a member of technical staff in the Teletraffic Theory and System Performance Department. His current research interests are performance analysis of packetized voice/data/video networks and the broadband Integrated Services Digital Network*
(continued on page 71)

One of several congestion-control methods in integrated packet networks is a bit-dropping scheme in which voice is encoded, using the embedded coding principle, and packetized with blocks of bits organized according to their order of significance. In a state of overload, the less significant blocks are dropped in the multiplexer queue to relieve congestion. Fluctuations in voice quality occur due to dynamically varying bit rate during a call. In this paper, we present methodologies for evaluating the performance of variable-bit-rate voice under fixed and variable loads. We have developed a Markov chain model to describe the probabilistic bit-dropping pattern corresponding to any specified traffic conditions. The model is used with a software tool to emulate packetized voice under various loading conditions. The resulting bit-dropped voice packet stream is then decoded, converted to an analog signal, and presented to listeners in a subjective test. We conclude that, with prudent traffic engineering of the network, voice quality will remain robust to temporal variations in bits per sample.

Introduction

To provide a broader range of services in a more cost-effective manner, it is necessary to consider network architectures that are integrated from both a customer and an internal network perspective.[1] Instead of transmitting voice, data, and video information over multiple specialized networks, a single integrated network can be used to support a wide range of services over a common set of facilities and through a common set of switching devices. An integrated network architecture should provide more efficient operations, administration, and maintenance and should allow new services not possible with

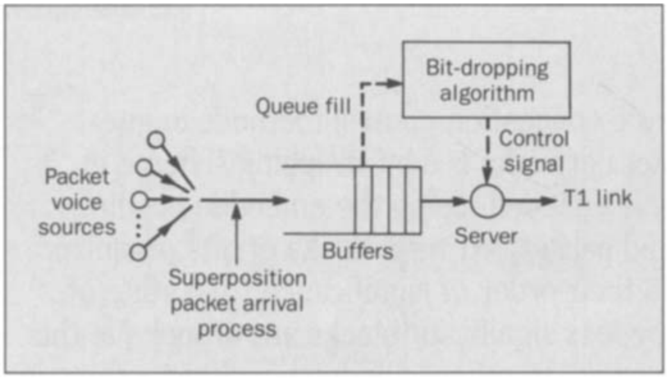

Figure 1. Packet voice multiplexer with bit dropping.

nonintegrated multiple networks.

To address the need for integrated voice-data-image networks required by the advanced phases of Universal Information Services, AT&T has been evaluating wideband packet technology (WPT).[2] In WPT, information is transported via a packet protocol. End-to-end connections are no longer circuit-oriented but are "virtual," in the sense that they exist only while information is being transmitted actively. This, in turn, leads to a more efficient use of bandwidth. Voice packets, for example, are generated only during periods of active speech and not during silent intervals.

Packet-based data networks have been commercially available for many years [e.g., the Advanced Research Projects Administration Network (ARPANET) and AT&T's Accunet® packet service]. As a result, the performance issues associated with packetized data communications are generally well understood. Packetized voice communication, however, required new studies for understanding the effect of randomness in traffic on voice quality. Because continuous speech is made discrete through packetization, a packet network must be designed to restore continuity of speech. By using a build-out delay equalization strategy at the depacketizer, we can force a constant delay for all voice packets, thereby eliminating temporal discontinuities in the speech.

Another performance issue associated with packetized voice communications is the method used to deal with congestion in the packet network. In WPT, a voice packet bit-dropping strategy is employed to control congestion. The bit-dropping scheme, based on embedded adaptive differential pulse-code modulation (ADPCM) voice coding, allows for overload control at any point in the packet network.[3-6] Subjective voice-quality evaluations have demonstrated that the bit-dropping strategy significantly outperforms more traditional packet-dropping strategies under WPT congestion.

Dropping bits to control congestion has one important consequence generally not found in standard circuit-based network architectures: *voice transmission quality dynamically varies with overall network loading.* As load increases in the packet network, the quantization noise also increases, resulting in reduced speech quality. If the network is not designed properly, voice quality could vary noticeably during the course of a single conversation. This paper reports our attempts to capture this effect through several interrelated approaches. We conclude that, with prudent traffic engineering of the network, voice quality will remain robust in the presence of temporal variations in bits per sample resulting from nominal load fluctuations.

A Packet Multiplexer with Bit Dropping

A schematic of a packet voice multiplexer is shown in Figure 1. Each voice source is packetized, and the superposition of all sources constitutes the packet arrival process at the queue in the multiplexer. In general, the multiplexer would also serve various data sources. We assume that the queue is finite with a buffer capacity of K packets and that the service discipline is first-in, first-out. During a 16-millisecond (ms) interval, 128 samples are collected and transcoded, using an embedded ADPCM scheme at a rate of 32 kilobits per second (kb/s). The speech samples are organized into a packet of four blocks. All the least significant bits from the 128 samples are contained in block 1 of the packet, the next most significant bits in block 2, and the two most significant in blocks 3 and 4. A 10-byte header incorporates a range of information about the packet, such as its destination, time-stamp information for build-out at the receiver, and other protocol-related information. Silence detection is employed so that voice packets are generated only during talkspurt periods.

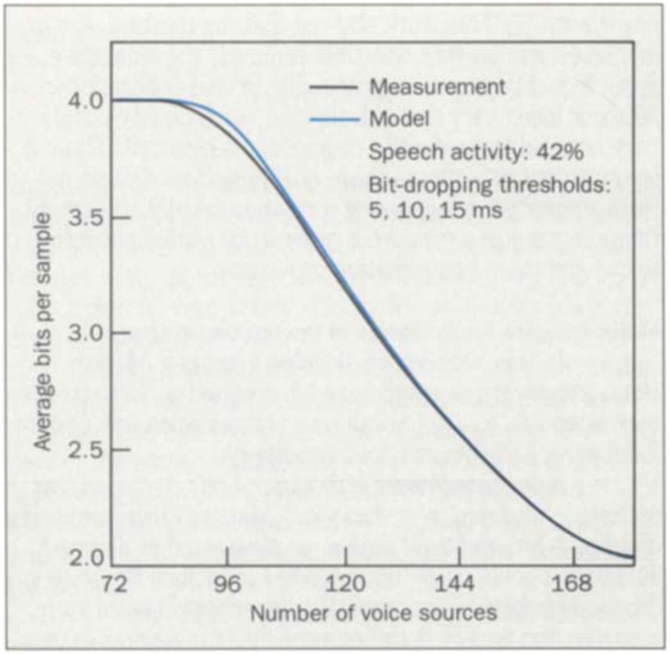

Figure 2. Average bits per sample as a function of load.

Silence periods are automatically restored at the receiver as part of the play-out strategy by using time-stamp information in the packet headers.

A bit-dropping algorithm is used in the multiplexer. Let L denote the current queue fill in packets. The bit dropping on the packets is done at the output of the queue—i.e., at the server—just before transmission (Figure 1). When L is smaller than the first threshold, no bit dropping is required and the voice packets have all 4 bits per sample. When L exceeds the first threshold Q_1, but is still smaller than a second threshold Q_2, block 1 containing the least significant bits is dropped. When L exceeds Q_2, blocks 1 and 2 are both dropped, thus reducing the information in the packet to the two most significant bits. The transmitted voice packet sizes are 74, 58, and 42 bytes, corresponding to the three states of the queue; the corresponding service times (on a T1 link) are 0.385, 0.302, and 0.219 ms. A packet is lost if the queue is full when it arrives.

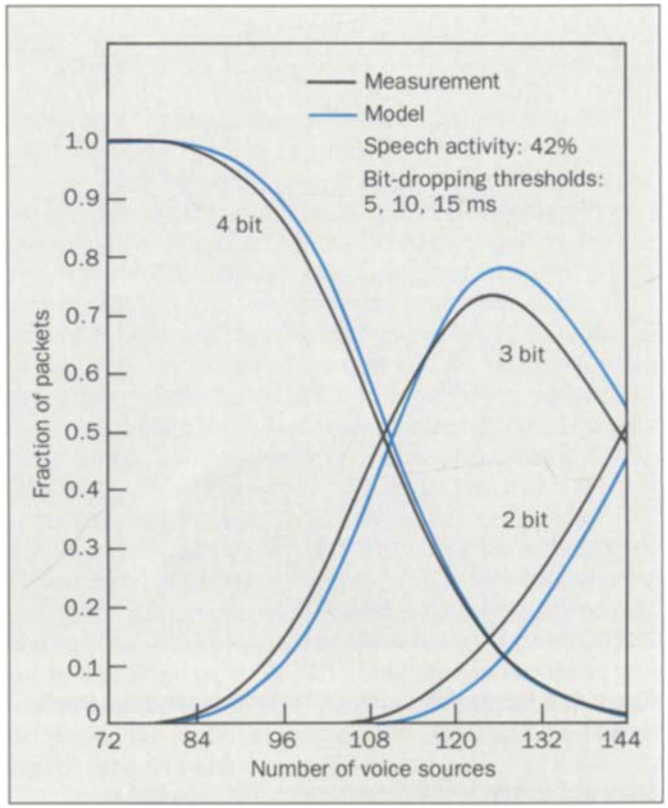

Figure 3. Fractions of packets with 4, 3, and 2 bits per sample.

Recent studies have shown that superposition of packet sequences generated by packetized voice sources with speech detection produces high burstiness (relative to a Poisson process) due to inherent correlations between successive interarrival times in the superposition stream.[7,8] In a packet multiplexer without bit dropping, these correlations tend to cause significantly larger queueing delays and packet losses than would be predicted by a Poisson model. Bit dropping on voice packets significantly smooths the superposition packet voice process by speeding up the packet service rate during critical periods of congestion in the queue, and the superposition packet arrivals can be viewed as a Poisson process from a packet

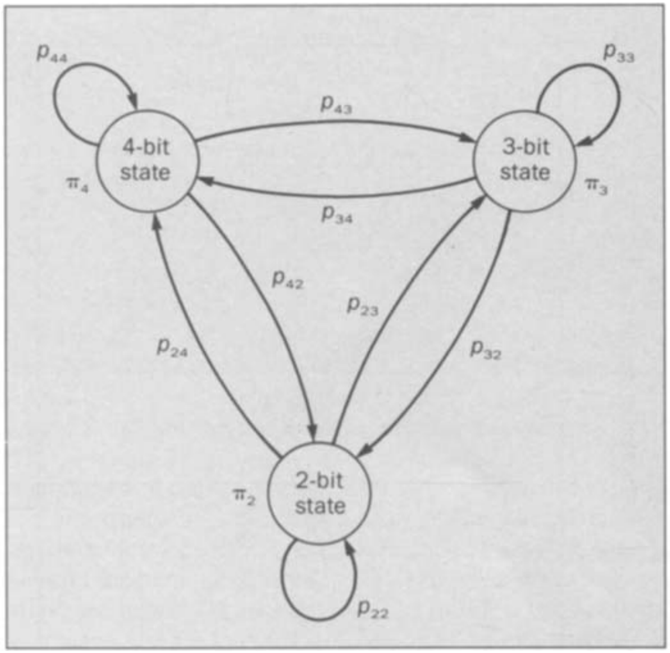

Figure 4. A Markov chain model for bit dropping at a fixed load.

delay and queue length perspective.[9] Sriram and Lucantoni[9] modeled the multiplexer as a queueing system with a state-dependent server whose service speeds up during states of congestion. The specific model used was an $M/\tilde{D}/1/K$ model, where $\tilde{D}$ denotes the deterministic but state-dependent nature of service. The model was used to compute various quantities of interest for performance characterization (e.g., carried mean bit rate for a voice call, queue length distribution, packet loss due to queue overflows, and mean and variance of voice packet delay). The predictions by the model are quite close to the measurements made with a hardware implementation of the multiplexer, as shown in Figures 2 and 3. Figure 2 shows the average bits per sample. Figure 3 shows the distribution of packets with 4, 3, and 2 bits per sample.

The subjective effects of bit dropping on voice quality under fixed network loadings have been reported elsewhere.[10,11] This work showed that, as the load was increased and bit rate was thus reduced, the subjective speech quality degraded gracefully. In real applications, network loads vary dynamically and voice quality could vary noticeably during the course of a single call. Thus, it is of interest to evaluate voice quality under realistic network scenarios, where loads vary dynamically and the bit rate allocated to a particular voice trunk varies according to call arrival and termination processes.

Methodologies for Synthesis of Packet Sequences

In this section, we develop a general Markov chain framework to synthesize bit-dropped packet streams corresponding to a voice call on a transmission link under fixed as well as dynamic load conditions.

Bit-Dropping Model. Dropping of bits in successive packets is modeled by a stationary Markov chain consisting of 4-bit, 3-bit, and 2-bit states, as illustrated in Figure 4. The model could easily be extended to include the case of dropped packets by incorporating a packet-dropped state; however, the packet-dropping probability is negligible over the range of load that is of interest.[9] The transition probabilities from state i to state j are denoted by p_{ij}. The packet state probabilities are denoted by π_i, where the subscript represents the number of bits per sample in the packet. These probabilities are the same as the fraction of 4-, 3-, and 2-bit packets that is realized for a given average bits per sample, say $\bar{b}$ (see Figures 2 and 3). In general, a given average bits per sample can be realized with various 4-, 3-, and 2-bit packet combinations, and the different fractions result in different variances for a specified $\bar{b}$.

Let $\Pi = \{\pi_i;\ i = 4, 3, 2\}$ denote the state probability vector and $P = [p_{ij};\ i,j = 4, 3, 2]$ denote the transition probability matrix. Note that

$$\sum_{i=2}^{4} \pi_i = 1 \tag{1}$$

$$\sum_{j=2}^{4} p_{ij} = 1 \qquad \text{for } i = 2, 3, 4 \tag{2}$$

Under steady-state conditions for a stationary Markov chain, we have

$$\Pi P = \Pi \tag{3}$$

Equations (1), (2), and (3) provide a set of relations indicating that, given two of the state probabilities, the three-state Markov chain is completely characterized if certain 4-tuples of the p_{ij} are specified. For example, π_4, π_3, p_{44}, p_{33}, p_{22}, and p_{43} completely specify the stationary Markov chain of Figure 4.

Both modeling and measurement data provide the fractions of 4-bit, 3-bit, and 2-bit packets that result in a given average bits per sample (see Figures 2 and 3). Hence, the state probabilities, π_i, are known for a specified average bits per sample. Also, if one specifies the desired mean and variance, the 4-, 3-, and 2-bit packet fractions can be determined; hence, the state probabilities are known.

While the state probabilities determine the average bits per sample, $\bar{b}$, and the associated overall variance, it is the transition probabilities that characterize the temporal (short-term) variations in the bits per sample. The queue dynamics of the packet multiplexer can be such that, over a call holding time interval $(0, T)$, two different cases of temporal variations can have the same average bits per sample and overall variance. However, over shorter segments of time within the interval $(0, T)$, the variance in bits per sample can vary over consecutive segments. This effect is characterized by the transition probabilities.

Link transition probabilities. In the stationary Markov chain model for bit dropping on successive packets, the state transition probabilities can be computed approximately by assuming exponential distributions for the corresponding queue state transition times (i.e., durations of queue fill remaining above/below the respective bit-dropping thresholds). If μ denotes the mean packet multiplexer service interval per packet, then $e^{-\mu/T}$ is the probability that no transition occurs during the interval of length μ, where T is the mean time to a transition for the corresponding queue state. Therefore, the probability that a transition occurs during a service interval is given by

$$p_{ij} = 1 - e^{-\mu_i/\bar{t}_{ij}} \tag{4}$$

where $\bar{t}_{ij}$ is the mean time to a transition from state i to state j and μ_i is dependent on the state i.

For a two-state system, the average bits per sample and one of the p_{ij} completely specify the Markov chain. If the case of transitions from 4-bit state to 2-bit state or vice versa is not considered for a single packet multiplexer, then $p_{42} = p_{24} = 0$. This is a reasonable assumption, since it is highly unlikely that the queue length will change by the difference between the first and second bit-dropping thresholds during the service time interval of a packet. In this case, for the three-state system, the average bits per sample, one of the π_i, and any 2-tuples of p_{ij} (except for 2-tuples from either the set p_{32}, p_{23}, p_{22} or the set p_{43}, p_{34}, p_{44}) completely determine the Markov chain probabilities.

The modeling/simulation studies offer estimates of $\bar{t}_{ij}$. For a two-state system with $\bar{b} = 3.7$, the studies indicate that $\bar{t}_{43}$ is about 5 ms. The transition probabilities can be computed using equations (1) to (4). In equation (4), $\mu_4 = 0.385$ and $\mu_3 = 0.302$. Using the Markov chain model with the computed transition probabilities, we can determine the profile of average bits per sample as a function of time on a link.

Fixed Load. At a fixed load, the correlations from packet to packet for a voice call are negligible under steady-state conditions, because the queue dynamics of a packet multiplexer stabilize within a few multiples of the packet service time.[9] The packet service time is of the order of 0.385 ms for a 4-bit packet, while the 16-ms packet generation interval of an individual voice call is much longer. Thus, we can assume that the state of a packet (that is, bit dropping experienced by a packet) is independent of the state of the previous packet for a voice call. This independence assumption suggests that a single voice call may be modeled as a chain of independent state occurrences ($p_{ij} = \pi_j$ in Figure 4). Consider the 3-bit state as the current state. The probability of the next state being 4-bit is π_4, being 3-bit is π_3, and being 2-bit is π_2. That is, whatever the current state, the transition probability of the next state being the i-bit state ($i = 2, 3, 4$) is

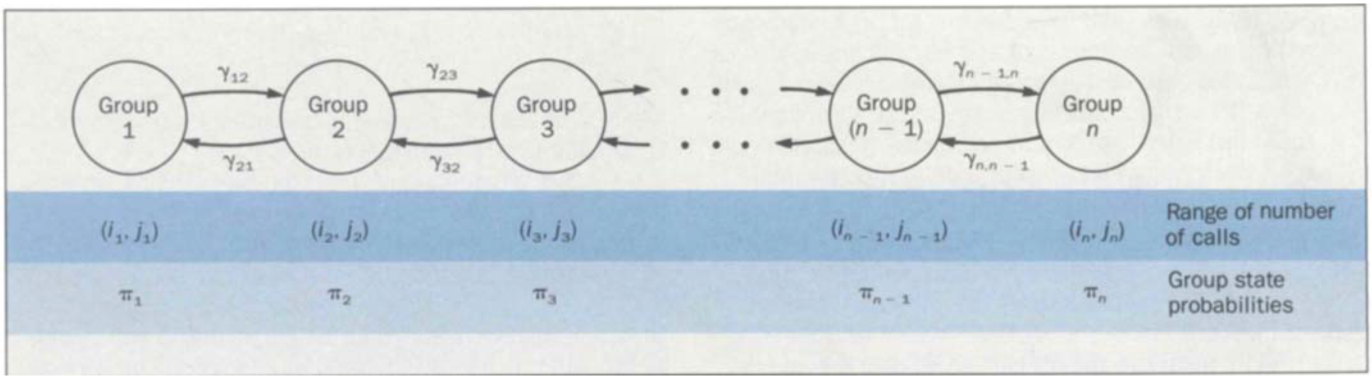

Figure 5. A Markov chain model for voice calls under dynamic load.

the same as the steady-state probability of the i-bit state. This captures the assumption that the state of a voice packet is independent of the state of the previous packet of the same voice call. The model now may be viewed as a fictitious/degenerate Markov chain. Hence, the proposed voice call model fits within the framework of the bit-dropping model outlined earlier. In addition, equations (1) to (3) are applicable, with the state transition matrix P given by $p_{ij} = \pi_j$; $i,j = 4, 3, 2$.

Thus, while correlations exist for bit dropping on successive packets of the aggregate packet stream from all calls on a link, the independence assumption holds well for packets belonging to a particular call. The average bits per sample and the fractions of packets with 4, 3, and 2 bits per sample are therefore sufficient to emulate bit-dropping effects on the packets of an individual voice call at a fixed load.

Dynamic Load. In most applications, the number of active voice calls on a link fluctuates during the course of a voice call. In these applications, packet-to-packet correlations become significant and the duration of these correlations is comparable to the time scales involved in the load fluctuations. The independence assumption (regarding bit dropping from one packet to the next) is no longer valid. Load fluctuations, however, can be accounted for by quantizing the number of active channels into various bins and modeling the load variations as yet another Markov chain. Consider Figure 5, which shows an n-state discrete Markov chain. Let i_k and j_k, respectively, denote the lowest and highest number of calls in group state k. The state probabilities for group k, Π_k, can be determined by noting that they are the sum of the probabilities of i_k to j_k channels being active. For Poisson call arrivals, exponential holding times, and blocked calls cleared, these probabilities can be computed using the following Erlang B formula:

$$p_l = \frac{a^l/l!}{\sum_{j=0}^{M} a^j/j!} \tag{5}$$

where p_l denotes the probability of l channels being busy, a is the offered load, and M is the maximum number of channels available. Then we have

$$\Pi_k = \sum_{l=i_k}^{j_k} \tag{6}$$

Strictly speaking, the p_l in equation (5) should be modified by the factor $1 / (1 - p_0)$ because the case of zero calls is not applicable. However, p_0 is very small.

The state probabilities, Π_k, can be computed using equations (5) and (6). For each state, the mean bits per sample, $\bar{b}_k$, is determined as the mean of the average bits per sample (which can be obtained from Figure 2) for i_k to j_k active voice calls. Then, the overall average bits per sample realized on a voice call under dynamic load is

$$\bar{b} = \sum_{k=1}^{n} \Pi_k \bar{b}_k \tag{7}$$

where n is the number of call states of the discrete Markov chain in Figure 5. Each state of the Markov chain is characterized by a different set of state probabilities ($\pi_4^{(k)}$, $\pi_3^{(k)}$, $\pi_2^{(k)}$, which can be obtained from Figures 2 and 3). Each set corresponds to the $\bar{b}_k$ realizable for that group of active calls.

The discrete Markov chain is completely characterized if the transition probabilities are also specified. As shown in Appendix A, the transitions from state k to $k + 1$ or to $k - 1$ may be modeled exactly by a phase-type distribution. However, the transition times can be approximated by a suitable, simple distribution (e.g., exponential or hyperexponential) whose mean and variance are matched to the phase-type distribution (see Appendix A for details). With the state probabilities, Π_k ($k = 1, 2, \ldots, n$), determined by equations (5) and (6), and the transition times, γ_{ij}s, determined by the method described in Appendix A, the discrete Markov chain of Figure 5 is completely characterized (γ_{ij} denotes the time spent in group state i before a transition is made to group state j). Packet sequences for a voice call under dynamic load can be synthesized using the above characterization.

While the above methodology of generating packet sequences for a voice call on a link that is experiencing varying load is useful, it is feasible to achieve the same objective by an approximate, but reasonably accurate and simple, technique:

1. Generate the number of active calls on a link as a function of time over an interval T (the average holding time of a voice call) using a simulation model with Poisson call arrivals and exponential holding time.
2. Map the sample path thus generated into quantized states, resulting in a sequence of subintervals $\{\tau_k^{(i)}\}$, $i = 1, 2, \ldots, m$, where the subscript denotes the quantized state and the superscript refers to the time subinterval. The sequence is such that

$$\sum_{i=1}^{m} \tau_k^{(i)} = T \tag{8}$$

and the overall average bits per sample is given by

$$\bar{b} = \frac{1}{T} \sum_{i=1}^{m} \tau_k^{(i)} \bar{b}_k \tag{9}$$

where $\bar{b}_k$ denotes the mean bits per sample for state k.
3. Use the packet sequence generation algorithm for a voice call under fixed load to generate the packet stream corresponding to $\bar{b}_k$ bits per sample for $\tau_k^{(i)}$ seconds.
4. Concatenate the above packet streams to synthesize the complete packet stream for the T-second voice call under consideration.

If the quantization of the states in the packet synthesis procedure is too coarse, the average bits per sample realized on the voice calls might not be a smooth function over consecutive time subintervals. If the quantization of the states is too fine—that is, if $\tau_k^{(i)}$ are too small (say, a second or two)—it may not be possible to generate meaningful packet streams. The simulation results have indicated that the quantization step size for creation of the discrete Markov chain of Figure 5 should be about three to six calls. The number of states in the discrete Markov chain need not be large, since voice calls receive about 4 bits per sample for up to about 90 active calls on the link (see Figure 2). When the maximum number of calls is 120, the discrete Markov chain model would contain five states, each with about six calls, in addition to the first state, which represents 1 to 90 active calls.

Figure 6 shows the number of calls in progress, generated using a model with Poisson call arrivals and exponential holding times (the average number of calls is 103). It also shows the corresponding average bits per sample realized on a voice call, generated using the approximate packet synthesis technique (the average bit rate is 3.7). It is clear that $\bar{b}$ closely tracks the number of active calls on the link.

The techniques developed here also apply to studies of the effect of peakedness in traffic on voice calls. For the discrete Markov chain approach, the p_l in equation (6) can be determined by using the equivalent random method.[12] For the approximate method, the sample path

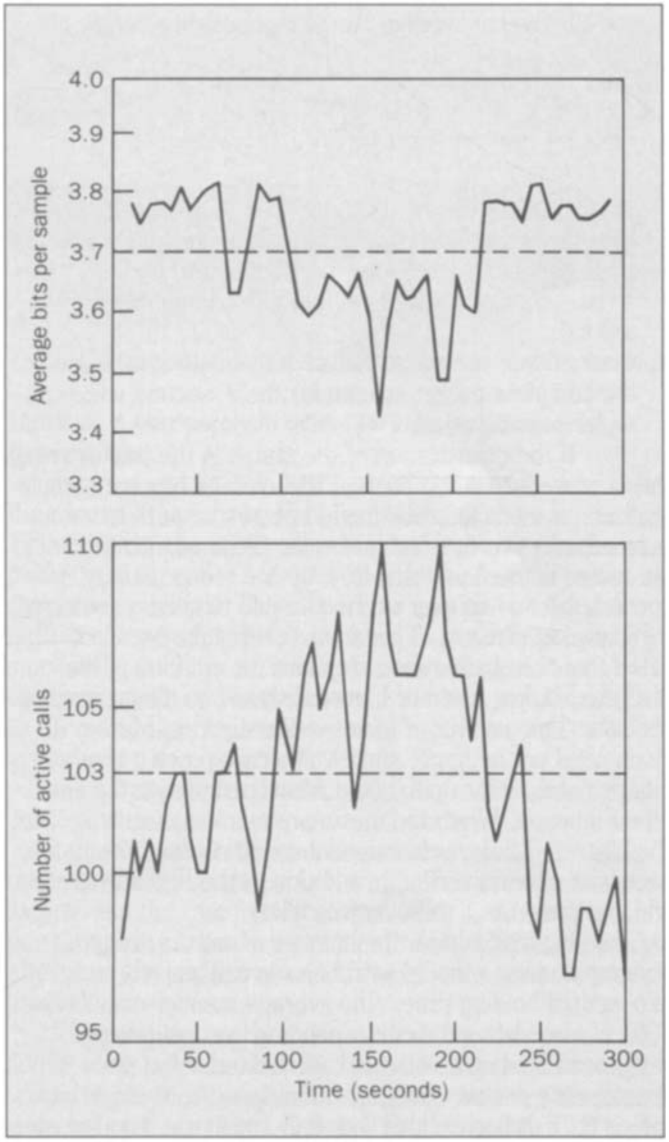

Figure 6. Variations in average bits per sample with time for a voice call (top) and the corresponding number of active calls on the link over the same time interval (bottom).

generated in the first step should include the peakedness in traffic. In integrated packet networks, it is usually the case that voiceband data calls are also served by the same multiplexer queue. The presence of voiceband data can be handled approximately by a corresponding (virtual) increase in the number of active voice calls on the link, by observing that a voiceband data call is equivalent to about 2.5 voice calls from bandwidth considerations.

Subjective Measurements of Voice Quality

Using the models of the previous section, we can generate packet streams to emulate packetized voice under various loading conditions. In this section, we use the model-generated bit-dropping patterns in conjunction with a tool that allows software simulation of packetized speech, and we present the results of subjective measurements of voice quality.

Test Conditions. Three sets of conditions were tested in a formal subjective voice performance evaluation. The packet sequences corresponding to the test conditions were synthesized, using the models for a voice call under fixed as well as dynamic loads.

Fixed load conditions. This class of conditions attempted to characterize the performance of a voice call under fixed load. Ten different conditions were considered. The probabilities or fractions of 4-, 3-, and 2- bit packets for each $\bar{b}$ were those obtained via models/measurements (Figures 2 to 3). With these fractions as the transition probabilities in the state transition matrix of a Markov chain, packet sequences corresponding to the length of speech files were generated.

Different variance conditions. A fundamental issue is whether the specification of average bits per sample, $\bar{b}$, is sufficient to characterize the voice quality of a call in a packet network. The variance in bits per sample plays a potentially important role in determining the voice quality. The variance in bits per sample on a voice call is of two types: the overall variance on a call due to the distribution of 4-, 3-, and 2- bit packets, and the variance due to short-term (in relation to the duration of the call) fluctuations in average bits per sample. The first type depends on the distribution of packets with 4, 3, and 2 bits per sample, whereas the second type depends on the time fluctuations in bits per sample. Here, we will consider the case of overall variance due to different packet distributions on a voice call.

Four conditions—(π_4, π_3, π_2) = (0, 1, 0), (.11, .78, .11), (.22, .56, .22), and (1/3, 1/3, 1/3), each resulting in an average of 3.0 bits per sample—were used for studying the effect of variance on voice quality. The first condition constitutes a reference case with all-3-bit packets. The second condition represents the expected distribution of packets out of a packet multiplexer (see Figure 3). The third condition is feasible on a link in a packet network (say, due to random merging/splitting traffic patterns at nodes internal to a packet network). The last condition reflects the case of equal numbers of 4-, 3-, and 2-bit packets. With the above fractions as the transition probabilities in the Markov chain model for a call under fixed load, voice packet sequences were generated.

Dynamic load conditions. This class of conditions attempted to characterize the performance of a voice call under dynamic load. The six conditions considered had the basic constraint that a voice call receive an overall average of 3.7 bits per sample.

In the first four conditions, the length of speech files was partitioned into five equal segments, each segment receiving different average bits per sample. The average bits per sample for each segment was determined so that the overall average is 3.7 bits per sample and the overall variance is about the same. These variations represent a realistic case. The mean bits per sample profiles for the four test conditions generated with the above criteria (good-to-bad, bad-to-good, good-bad-good, and bad-good-bad) are illustrated in Figure 7. A fixed load condition with an average of 3.7 bits per sample is also shown in the figure.

The last two conditions, in which the bit rate varied over the course of the speech segment according to a Poisson call model, attempted to characterize the performance of a voice call on a T1 digital facility [1.536 megabits per second (Mb/s)]. From the standard Poisson call arrival model with an offered load of 104 erlangs and the assumption that call duration is exponentially distrib-

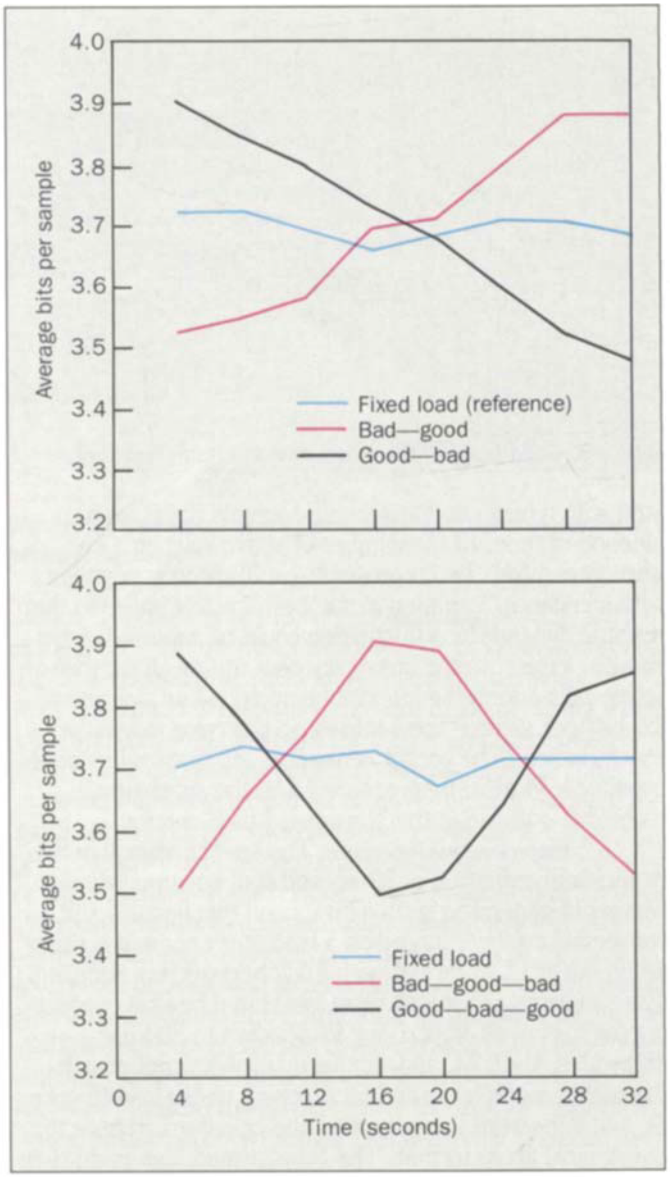

Figure 7. Variations in average bits per sample over time for some test conditions. (Overall average bits per sample = 3.7.)

65

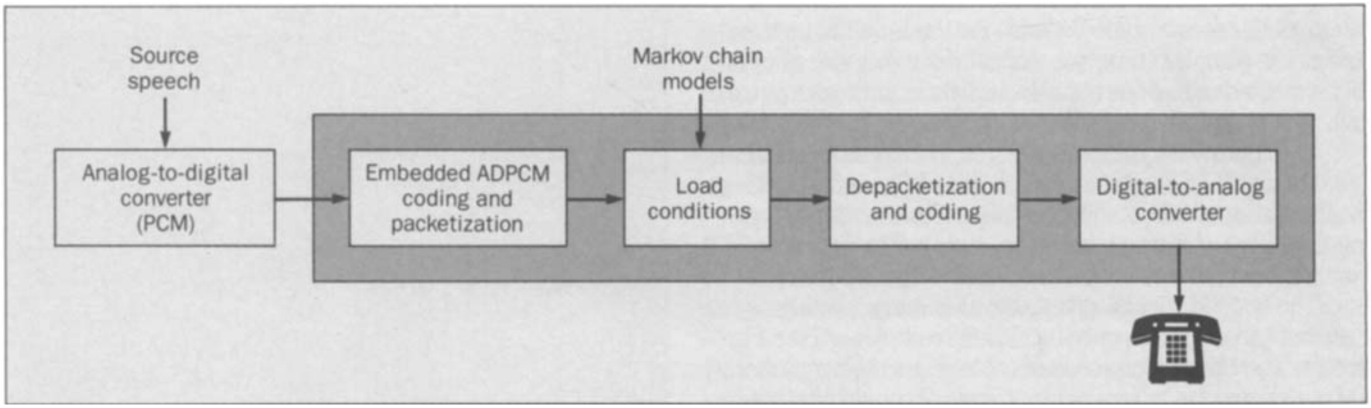

Figure 8. Schematic of the voice packet simulation tool.

uted with a mean of 300 seconds, sample paths, as a function of time, of the number of active calls on a link were generated. Two representative 32-second segments with substantial variation in number of active calls (so that realistic dynamic load fluctuation could be captured) were chosen, subject to the constraint that the mean number of active calls during the interval be such that an average of 3.7 bits per sample was realized on the voice call under consideration. The packet sequences for voice calls under dynamic load were then generated by the approximate technique outlined in the "Dynamic Load" section.

Subjective Test Procedure. The speech stimuli in the experiment consisted of 32-second conversational speech segments generated by two males and two females. All conversations were stored on a laboratory computer using 16-bit linear PCM format with 8-kilohertz (kHz) sampling. The simulated conditions were generated by first converting the files to 64-kb/s μ-law PCM, then to 32-kb/s embedded ADPCM, and simulating packet production by the packet multiplexer and bit dropping under load (Figure 8). The files were then decoded and transformed back to the original linear format. The transformed files could then be played out of a digital-to-analog (D/A) converter, through a simulated telephone connection, to the experimental participants.

Forty-four adults, recruited outside AT&T, participated in a formal subjective listening test. The participants were asked to rate the quality of speech segments along a five-point quality scale: excellent, good, fair, poor, unsatisfactory. Following standard practice, the mean opinion scores (MOS) for each condition were derived from the voice quality ratings by first transforming the ratings to a numeric scale (1 = unsatisfactory, 2 = poor, 3 = fair, 4 = good, 5 = excellent) and then taking the arithmetic averages of the transformed values.

Results. Mean opinion scores are plotted in Figure 9 for the fixed load conditions in which the average bit rate was held relatively constant over the course of the speech segments. The open circles represent MOS measurements for the simulated packet stream with fixed average bits per sample. Plotted on the same graph are MOS values (solid diamonds) from the earlier listening test in which subjects rated the quality of speech recorded through prototype packet multiplexer hardware. Aside from a small offset due to hardware-related factors (e.g., channel unit noise), there appears to be good agreement between the two tests for conditions in which the load is fixed.

The effect of different variances on speech quality, with the average bit rate held constant at 3.0 bits per sample, is depicted in Figure 10. The data show clearly that, as the variance in the speech segment increases, the overall quality of that segment decreases. An analysis of variance of the test data demonstrated this effect to be highly significant.

MOS ratings for the six different conditions in which the short-term average bits per sample was free to

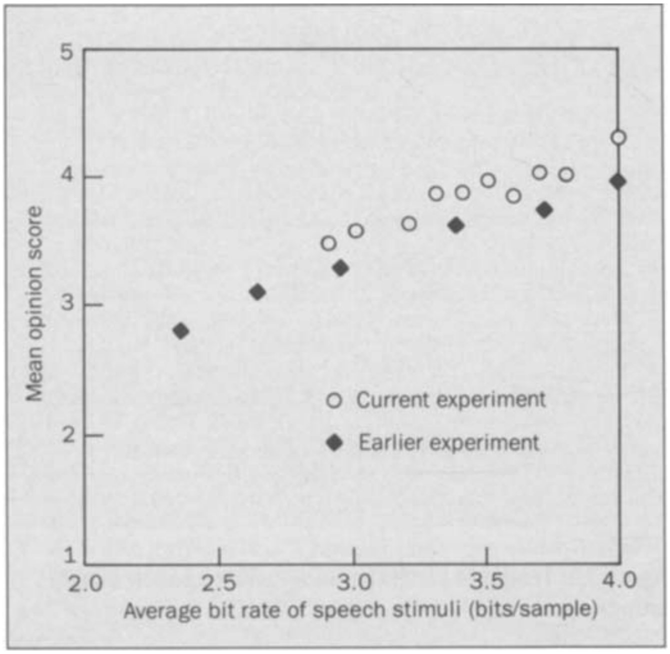

Figure 9. Mean opinion scores for different average bit rates at fixed loads.

vary dynamically about a long-term average of 3.7 bits per sample are shown in Figure 11. Also shown is the rating obtained for an average of 3.7 bits per sample for a voice call under fixed load. Interesting, though minor, differences were observed in the MOS values obtained among some of these seven conditions. The pattern of these differences (e.g., bad-to-good being rated better than good-to-bad) suggested that subjects had weighted their opinions of quality according to the distribution of quality within the 32-second interval. Seemingly, when the bit rate was higher at the end of the speech segment, subjects rated that condition better than when the bit rate was lower at the end of the speech segment.

We carried out an analysis in which we correlated the overall MOS values obtained for the 28 speech trials (7 conditions for 4 speakers) at an overall bit rate of 3.7 bits/sample with the short-term bit rates averaged within 2.64-second intervals across the length of the speech segments. The results of the correlational analysis are plotted in Figure 12, where the data points are at the midpoints of the 2.64-second intervals. They reveal the temporal weighting function that subjects were evidently using when judging the overall quality of the 32-second speech segments. The correlation between subjects' quality judgments and short-term bit rate was highest during the last 5 seconds of the speech segment. As the average bits per sample within this period increased, so did the quality ratings. The lowest correlations were found within approximately 8 seconds to approximately 18 seconds from the onset of the speech segments. The function plotted in Figure 12 can be regarded as an example of the serial-position effect, as described in the literature on human memory.[13] A multiple-regression analysis demonstrated that while the function plotted in Figure 12 is statistically reliable, it accounted for less than 2 percent of the overall variance in subjects' voice quality ratings, indicating that the effect is of relatively minor importance.

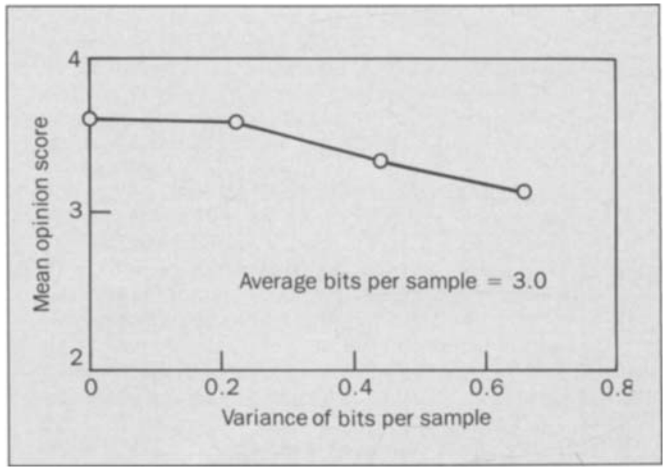

Figure 10. Mean opinion scores as a function of variance of bits per sample.

Conclusions and Discussion

In this paper, we have presented a methodology for synthesizing packet sequences for a voice call to emulate the effects of bit dropping (or variable bit rate) under a

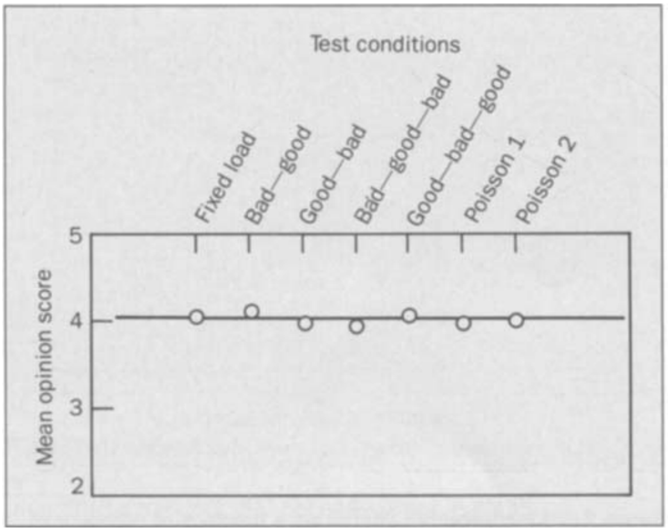

Figure 11. Sensitivity of mean opinion score to temporal variations in bits per sample.

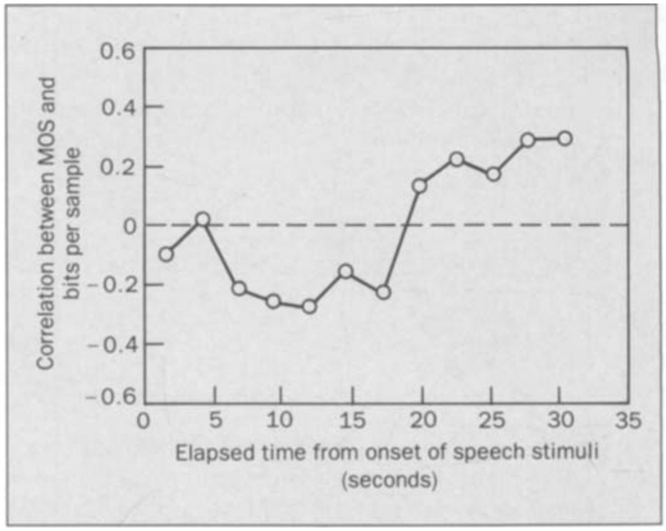

Figure 12. Temporal weighting function for speech quality ratings.

variety of load scenarios. The methodology incorporates temporal characteristics of bit dropping, such as the durations for which packets experience a certain level of bit dropping. Models for the emulation of the bit-dropped voice packet sequences have been described for fixed as well as dynamic voice load conditions. Using these models, we have characterized the subjective effects of variations in bit rate on voice quality.

Previous voice performance evaluations of packetized voice have examined the effects of reductions in the average bit rate on speech quality ratings[10,11] and on the quantization distortion introduced into the speech signal. In these evaluations, the bit rates varied according to statistics determined by the programmed settings of a voice load simulator. In the experiment reported here, bit rate variations were introduced under software control. This allowed us to evaluate speech quality under the influences of dynamic traffic fluctuations and the concomitant temporal changes in bit rate.

Results of subjective tests demonstrated that, in general, speech quality can be affected by variations in the bit rate over time, especially in a heavily loaded network when the mean bits per sample over a call duration is low (e.g., $\bar{b} = 3.0$). However, with prudent traffic engineering of the network, the mean bits per sample over a call duration can be kept fairly high (e.g., $\bar{b} = 3.7$). Then voice quality remains robust to temporal variations in bits per sample resulting from nominal load fluctuations.

The data presented in this paper extend our understanding of the transmission performance of packetized voice under dynamic load with temporal variations in bit rate within the duration of a call. Further, the models and methodology described facilitate the specification of performance objectives for variable-bit-rate voice in integrated packet networks.

Acknowledgments

We would like to thank C. A. Dvorak, who conceived the idea of the packet simulation software tool and was instrumental in making it a reality; the tool provided the motivation for model development and facilitated the subjective testing reported in this paper. Thanks are also due to L. Kofman for aiding in the software development and C. Ward for conducting the subjective tests.

References

1. J. L. Cummings, K. R. Hickey, and B. D. Kinney, "AT&T Network Architecture Evolution," *AT&T Technical Journal*, Vol. 66, No. 3, May-June 1987, pp. 2-12.
2. R. W. Muise, T. J. Schonfeld, and G. H. Zimmerman, "Experiments in Wideband Packet Technology," *Proceedings of Zurich Seminar on Digital Communication*, March 1986, pp. D4.1-D4.5.
3. D. J. Goodman, "Embedded DPCM for variable bit rate transmission," *IEEE Transactions on Communications*, COM-28, No. 7, July 1980.
4. N. Yin, S. Q. Li, and T. E. Stern, "Congestion Control for Packet Voice by Selective Packet Discarding," *Proceedings of Globecom*, '87, Tokyo, Vol. 3, November 1987, pp. 1782-1786.
5. K. Sato, H. Nakada, and Y. Sato, "Variable Rate Speech Coding and Network Delay Analysis for Universal Transport Network," *Proceedings of INFOCOM*, New Orleans, March 1988, pp. 0771-0780.
6. J. M. Holtzman, "The Interaction between Queueing and Voice Quality in Variable Bit Rate Packet Voice Systems," *Proceedings of International Telecommunications Conference*, Kyoto, Japan, 1985, Paper 2.2A-4.
7. K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986, pp. 833-846.
8. H. Heffes and D. M. Lucantoni, "A Markov-Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986, pp. 856-868.
9. K. Sriram and D. M. Lucantoni, "Traffic Smoothing Effects of Bit Dropping in a Packet Voice Multiplexer," *Proceedings of IEEE INFOCOM*, New Orleans, March 1988, pp. 0759-0770.
10. D. O. Bowker and C. A. Dvorak, "Speech Transmission Quality of Wideband Packet Technology," *Proceedings of Globecom* '87, Tokyo, Vol. 3, November 1987, pp. 1887-1889.
11. D. O. Bowker and C. B. Armitage, "Performance Issues for Packetized Voice Communications," *Proceedings of National Communications Forum*, Chicago, Vol. 41, No. 3, September 1987, pp. 1087-1092.
12. R. B. Cooper, *Introduction to Queueing Theory*, North Holland, New York, 1981.
13. R. L. Klatzky, *Human Memory: Structures and Processes*, W. H. Freedman and Company, San Francisco, 1975.
14. M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: Algorithmic Approach*, Johns Hopkins Press, Baltimore, 1981.

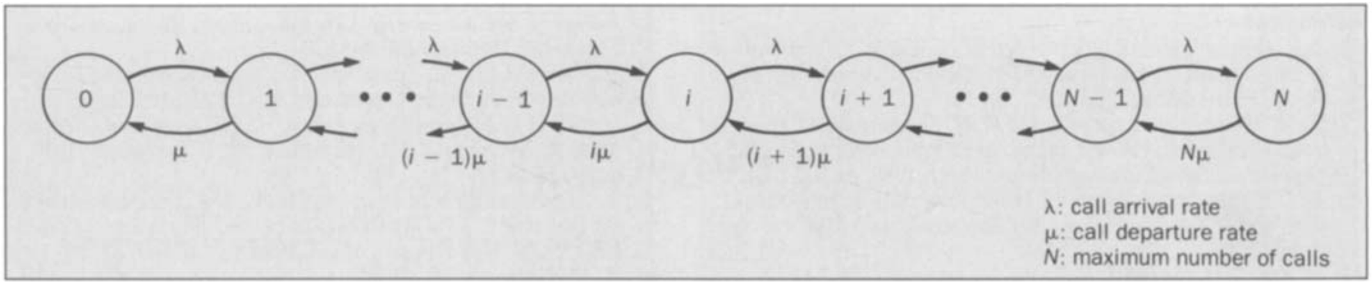

Figure A-1. A Markov chain representation of the call on/off process at a fixed offered load.

Appendix A. Modeling State Transitions

In a packet voice multiplexer, the number of active voice calls fluctuates according to a Markov chain, as shown in Figure A-1. As described in the section "Dynamic Load," we group the adjacent states to form a coarse representation of fluctuations in the number of active voice calls. As an example, consider the grouping of adjacent states to form three groups, shown in Figure A-2. Let the three groups of states be denoted by S_1, S_2, S_3:

$$\begin{aligned} S_1 &= (0, 1, 2, \ldots, j_1) \\ S_2 &= (i_2, i_2 + 1, \ldots, j_2) \\ S_3 &= (i_3, i_3 + 1, \ldots, j_3) \end{aligned} \tag{A-1}$$

Then the state transitions among the three groups are described by the following rate matrix:

$$R = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & A_{23} \\ 0 & A_{32} & A_{33} \end{bmatrix} \tag{A-2}$$

where A_{mn} is a matrix of dimension $(j_m - i_m + 1) \times (j_n - i_n + 1)$. This type of grouping aids us in simplifying the stochastic process used in simulation for the effects of call on/off variations on the bit-dropping process.

Our objective is to find the distribution of time spent within each group before a transition occurs to an adjacent group. It is well known in queueing theory[14] that for the formation given here, the time spent in any of the states S_i, $i = 1, 2, 3$, has a phase-type distribution because it is the time spent to absorption in a stationary Markov chain, which has one or more absorbing states. We can determine the mean and variance of the time spent in each group (S_1, S_2, S_3) until a transition to another group, conditioned on the group to which the transition occurs. The values of the means and the variances can be computed exactly by analysis. Then we can approximate the duration by a suitable, simple distribution (e.g., a hyper-exponential) whose mean and variance are matched to the values computed for the underlying phase-type distributions. This procedure facilitates the use of a simpler distribution, without losing much in the way of accuracy, thereby reducing the complexity of the simulator.

For illustration, consider a simple situation where group S_2 consists of only two states and so $j = i + 1$. The mean, β_1, and variance, β_2, of the phase-type distribution for time spent in group S_2 can be derived with some algebraic effort, and are as follows:

$$\beta_1 = \Delta^{-1}\left[(i+1)\mu + 2\lambda\right] \tag{A-3}$$

$$\beta_2 = 2\Delta^{-2}\left[3\lambda^2 + (5i+4)\lambda\mu + (i+1)^2\mu^2\right] \tag{A-4}$$

where

$$\Delta = \lambda^2 + i\lambda\mu + i(i+1)\mu^2 \tag{A-5}$$

The above quantities refer to the case when the sojourn through S_2 begins in state i and ends in state $i - 1$ or $i +$

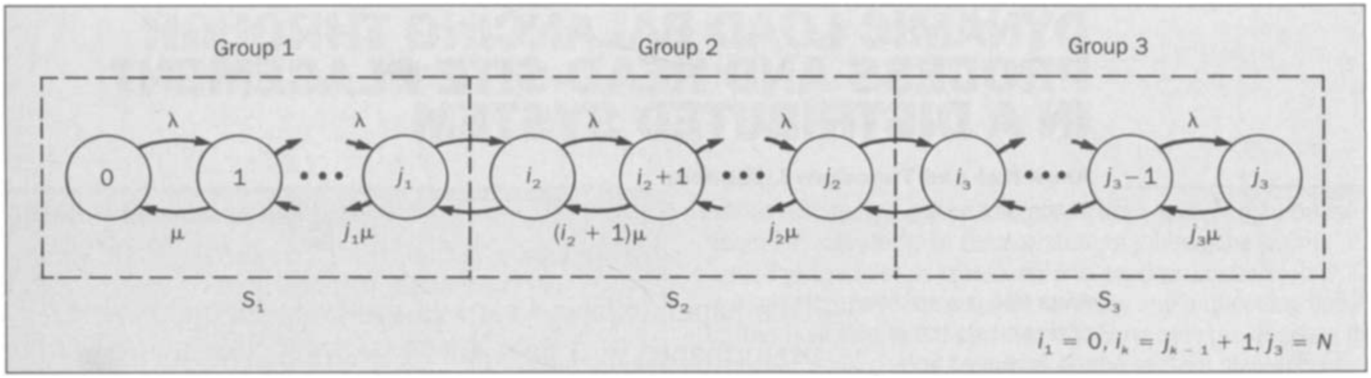

Figure A-2. Grouping of adjacent states to provide a coarse representation of call on/off variations.

2. The coefficient of variation of the sojourn time in S_2 is given by:

$$c^2 = \frac{\beta^2 - \beta_1^2}{\beta_1^2} = 1 + \frac{2(i - \rho)\rho}{4\rho^2 + 4(i + 1)\rho + (i + 1)^2} \qquad \text{(A-6)}$$

where $\rho = \lambda/\mu$ is the offered load in erlangs. To get a numerical feel for c^2 values, consider the following table:

	ρ (erlangs)	i	$j = i+1$	c^2
Case 1: $i, j < \rho$	104	97	98	0.9845
Case 2: $i, j > \rho$	104	109	110	1.0123

Note that the coefficient of variation of the time spent in group $S_2 = (i, i + 1)$ has a value less than 1 if the states in S_2 lie on the lower side of ρ, and has a value greater than 1 if the states lie on the higher side of ρ. In either case, the value of the coefficient of variation is close to 1, indicating the possibility of having an approximately exponential distribution.

The moments and coefficient of variation for the actual groups of states of interest can be computed exactly. Then suitable erlang, exponential, or hyperexponential approximations can be made, depending on whether $c^2 < 1$, $c^2 = 1$, or $c^2 > 1$. If, on the other hand, it is observed that c^2 is always quite close to 1, as seen in the simple example above, then the distributions of the times spent in the various groups of states may be approximated by exponential distributions with suitable rates. [Note that equation (A-6) suggests the possibility of having c^2 values close to 1 as long as the group of states i through j is such that i and j values are neither too small nor too large compared to ρ. Further, note that the probabilities are small that the Markov chain will be in states too far below or too far above ρ.]

As in the section "Dynamic Load," let γ_{ij} denote the time spent in group S_i (group i in Figure 5) before making a transition to group S_j. Clearly, γ_{ij}s are also phase-type distributed. Once the appropriate absorbing state and the transition rates are identified, the mean and variance of the conditional transition times, γ_{ij}, can be computed using the same procedure as described above.

Biographies (continued)

(ISDN). He received B.S. and M.S. degrees from the Indian Institute of Technology, Kanpur, and a Ph.D. from Syracuse University, all in electrical engineering. He joined AT&T in 1983. Mr. Bowker is a distinguished member of technical staff in the Network Performance Planning Department. He conducts voice performance evaluations of transmission devices and digital speech coding algorithms, with emphasis on performance under realistic network scenarios. He has a B.S. from the State University of New York at Brockport and M.S. and Ph.D. degrees in experimental psychology from the University of Rochester. He joined AT&T in 1980.

(Manuscript received August 26, 1988)