Analysis of Multipoint Videoconferencing under Reroutable Route-Configuration Assignment

Tat-Keung Chan and Tak-Shing Peter Yum*
Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong

In this paper, we study the use of reroutable assignment for multipoint videoconferences in a high-speed network. A conference model is constructed and conference calls are classified. A conference of a particular type can ride on different route-configurations. According to the location of the current speaker, a conference has different modes of operation. Two network management functions are discussed: call admission ensures a preset quality-of-service requirement by blocking new calls that causes congestion; route-configuration assignment determines the multicast tree for distributing the video of the current speaker. The reroutable route-configuration assignment is introduced. It allows a change of route-configuration when there is a change of speaker. Two reroutable assignment schemes are studied. In the normal scheme, a conference is always rerouted to the least congested route-configuration; while in the sticky scheme, a conference is only rerouted when the current route-configuration is congested. The video freeze probability, rerouting probability and the extended capacity space are derived. An example shows that the video freeze probabilities of the two schemes do not differ significantly. The sticky scheme, however, is superior as it gives a much smaller rerouting probability than the normal scheme. © 1998 John Wiley & Sons, Inc.

1. INTRODUCTION

   Multimedia is identified to be a major trend in communication and computing. Among the numerous kinds of multimedia broadband services, videoconferencing is predicted to be one of the most important for both business and residential users.[1,2] Videoconferencing systems allow a number of participants from different locations to exchange various types of information such as video, audio, and data. Many prototype desk-top conferencing systems have been built.[3-6] However, most of these prototype systems are built on local area networks. For large scale deployment of videoconferencing services over a wide

---

   * Author to whom correspondence should be addressed. e-mail: yum@ie.cuhk.edu.hk.

area network, special care and consideration should be given to video transmission because of its large bandwidth requirement.

There are various ways to present the video images of conferees at different locations and there are different ways of classifying video conferencing systems.[7-9] In this paper, we study a kind of multipoint videoconferencing system whereby only the video of the current speaker is transmitted to other conferees. This is not really restrictive as in a multiparty conference, speakers normally speak one at a time. In videoconferencing environment, the added advantage is that private sessions between any conferees can always be set up without affecting the main session. From a traffic engineering point of view these private sessions will be treated as separate conferences. The benefit of this restriction, however, is that videoconferencing can be conducted with a minimum amount of equipment and bandwidth. This kind of videoconferencing system requires a fixed multicast tree be set up when a call is initiated. When there is a change of speaker, say from A to B as shown in Figure 1, the transmission direction of the path between these two speakers has to be reversed. Statistical multiplexing such as the use of ATM technology allows a number of conferences to share a link and can improve the bandwidth utilization. However, statistical multiplexing means that with small probability the bandwidth of a reverse link might not be available when needed. When this occurs, video freeze will be experienced for certain conferees. We choose the video freeze probability as a measure of videoconferencing quality.

In Ref. 10, the conference performance under basic route-configuration assignment is studied. Basic route-configuration assignment means that the same route-configuration is maintained throughout the conference session. In this paper, we study the more sophisticated reroutable route-configuration
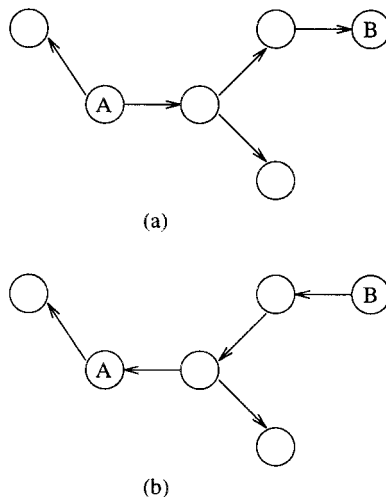

(a)

(b)

Figure 1. Path between old and new speaker is reversed, (a) A transmitting, (b) B transmitting.

assignment. Reroutable route-configuration assignment means that the route-configuration of the conference may change when there is a change of source node due to a change of speaker. Section 2 presents the conference network model. This includes the discussions of conference type, route-configuration and the mode of operation. In Section 3, we briefly describe the conference call admission and route-configuration assignment. We will introduce the basic assignment and the reroutable assignment with two specific reroutable schemes. Then in Section 4, the video freeze probability, the extended capacity, and the rerouting probability are derived. The call blocking probability is derived in section 5 and we conclude this paper in section 6.

2.   CONFERENCING NETWORK MODEL

The conferencing network model includes a few necessary assumptions similar to that in Ref. 11.

1. The speech duration of a conferee is an exponentially distributed random variable with mean μ^{-1}.
2. The transition from the current speaker to the next speaker is equally probable among all nonactive conferees.
3. In all conferences, voice traffic is unrestricted and hence in the case of a video freeze event, conference activities are not affected, only the conference quality is affected by the delayed onset of video when there is a change of speaker.
4. All videos are transmitted in the same format, and the capacity of each network link can be characterized in terms of the number of video channels (or the number of video transmissions it can carry simultaneously with guaranteed quality). The characterization is straightforward if dedicated circuits are set up as in a circuit switch network environment, or can be done via known queueing techniques[12,13] if videos are packetized and transmitted in store-and-forward fashion as in an ATM network environment. The queueing analysis in the latter case needs as inputs the calibrated video quality in terms of cell loss rate and delay statistics. It is possible to characterize videos transmitted in different formats. We do not address this issue here as it is more complicated.
5. Additional video communications, such as private point-to-point transmissions, are allowed whenever possible. However, they will be considered as different conference calls.

2.1.   Conference Type

Consider a particular conference on an N-node network. Let n_i be the number of conferees attached to node i and let $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ be conferee distribution. Nodes that have at least one conferee attached to, or $n_i > 0$, are called *conference nodes*. Then $w(\mathbf{n}) = \Sigma_{i=1}^{N} u_1(n_i)$ is the total number of conference nodes for the conference under consideration.† A conference node with the active conferee attached is denoted as the *source node*, other conference nodes are hence called the *sink nodes*. Let $n_{\mathrm{sum}} = \Sigma_{i=1}^{N} n_i$ be the total number of conferees in that conference. When the current speaker of a conference located at node i has finished speaking, the next speaker will either be one of the

---

† Where $u_1(n) \overset{\Delta}{=} \min(1, n)$ is an indicator function of whether $n > 0$ or not.
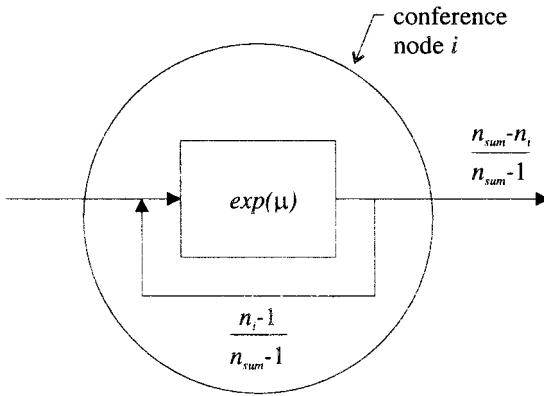
Figure 2. Active period of a conference node: a geometric sum of exponential random variables.

remaining $(n_i - 1)$ conferees at the same node with probability $[(n_i - 1)/(n_{\text{sum}} - 1)]$, or one of the conferees at the other nodes with probability $[(n_{\text{sum}} - n_i)/(n_{\text{sum}} - 1)]$. The active period of conference node i, denoted by T_i, is thus a geometric sum of exponential random variables as shown in Figure 2 and can be shown[14] to be exponentially distributed with mean $(n_{\text{sum}} - 1)/(n_{\text{sum}} - n_i)\mu$. For convenience, let

$$x_i = \begin{cases} \dfrac{(n_{\text{sum}} - n_i)}{(n_{\text{sum}} - 1)}\mu & n_i > 0 \\ 0 & n_i = 0 \end{cases} \tag{1}$$

such that $1/x_i$ is the mean active period of n_i provided that $n_i > 0$. Let $\mathbf{x}(\mathbf{n}) = (x_1, x_2, \ldots, x_N)$. Since we have assumed uniform transition probabilities among conferees, $\gamma_{i,j}$, the transition probability from conference node i to j is just

$$\gamma_{i,j} = \begin{cases} \dfrac{n_j}{n_{\text{sum}} - n_i} & j \neq i; n_j, n_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Depending on the distribution of the conference nodes, we have different *conference types* characterized by $\mathbf{x}(\mathbf{n})$. Let Z be the total number of conference types.

2.2. Route-Configuration

A conference of a particular type can be realized by a number of *route-configurations*. Each route-configuration is a set of bidirectional links connecting the set of conference nodes in the form of a tree. A route-configuration is said
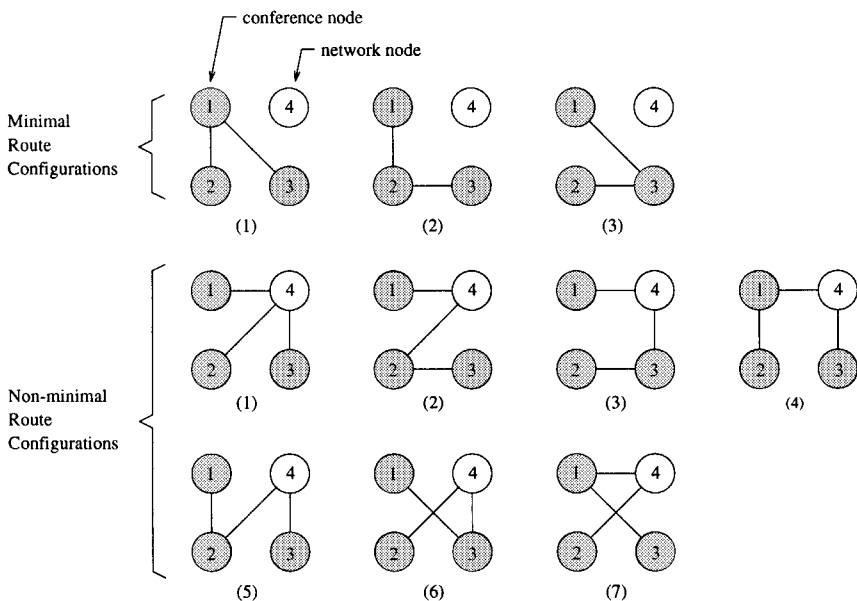
Figure 3. Route-configurations for a conference spanning nodes 1, 2, and 3 in a 4-node network.

to be *minimal* if it consists of only conference nodes and *nonminimal* otherwise. In the following, we shall restrict our study solely to the minimal route-configurations. Figure 3 shows the three minimal route-configurations and the seven nonminimal route-configurations for a conference type spanning nodes 1, 2, and 3 in a four-node network.

The total number of minimal route-configurations, $\alpha(m)$ for a conference spanning m nodes in a fully connected network can serve as an upper bound on the total number of minimal route-configurations for all networks. In graph theory terms, minimal-route configuration corresponds to a *labeled tree*. By Cayley's theorem (H. Prüfer, 1918), the number of labeled trees with m vertices is found to equal to $m^{(m-2)}$.[15] Table I shows the values of $\alpha(m)$ for m ranges from 3 to 8.

Sometimes, delay constraints may exclude configurations that have long hop counts between nodes and nodal processing power constraints may exclude configurations with large node degrees.

Table I. $\alpha(m)$ for $m = 3$ to 8.

m	3	4	5	6	7	8
$\alpha(m)$	3	16	125	1296	16,807	262,144

2.3.   Mode of Operation

A conference with $w(\mathbf{n})$ conference nodes has $w(\mathbf{n})$ *modes of operation* where each mode corresponds to one of the conference nodes being the source node. The mode of operation determines the link transmission direction in a particular-route configuration. Figure 4 shows the three modes of operation corresponding to a conference realized by minimal route-configuration (1) in Figure 3.

3.   CONFERENCE MANAGEMENT

We introduce in this section the network management functions needed for providing videoconferencing. Let there be a *control center* in the network responsible for call admissions and route-configuration assignments. It has global knowledge of the network status. For each conference, there is a *conference manager* located at one of the conference nodes. The conference manager is a computer process responsible for handling mode changes.

3.1.   Call Admission

Video freeze may occur for the kind of videoconferencing systems under consideration. To illustrate, Figure 5(a) shows a conference spanning four nodes on a particular route-configuration. Suppose there is a change of speaker from a conferee at node 2 to a conferee at node 4. If the same route-configuration is employed, the direction of the path between nodes 2 and 4 has to be reversed. Figure 5(b) shows the three possible cases for the conference to experience video freeze when such a change of speaker occurs. Note that link $(1, 3)$ is not blocked since it is established and used in the original route-configuration.

The probability of experiencing video freeze by any conferee in a conference is chosen as a measure of the quality of service (QoS) and the QoS is maintained by blocking conference calls that will cause unacceptably high video freeze probability. Thus when a new conference is initiated, a conference management process is created. The conference manager collects information
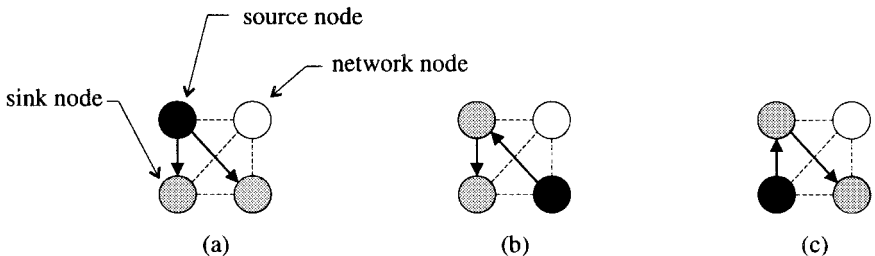

Figure 4.   The three modes of operation corresponding to minimal route-configuration (1) in Figure 3.
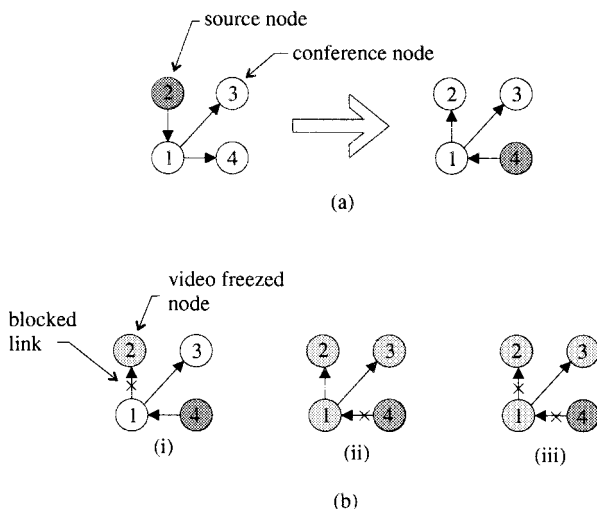
Figure 5.   (a) Change of modes, (b) three cases of video freeze.

such as the number of conferees, their locations, whether the conferees are ready for conferencing, etc. and sends a request to the control center. The control center checks the current bandwidth resources in the network and admits the conference call if the resulting traffic mix falls within the network capacity space; the conference call is rejected otherwise. The network capacity space will be derived in the next section.

3.2.   Basic Route-Configuration Assignment

When a new conference is admitted, it is assigned to one of the many possible route-configurations. With basic route-configuration assignment (basic assignment for short), a conference stays in its initial route-configuration throughout the conference session. Having selected a route-configuration for the new conference, the control center will inform the conference manager to establish the connections accordingly.

3.3.   Reroutable Route-Configuration Assignment

With the *reroutable* route-configuration assignment (reroutable assignment for short), the route-configuration of a conference is allowed to change, say, to a less congested route-configuration when there is a change of source node with the aim of reducing the video freeze probability. To illustrate the rerouting of conferences, Figure 6 shows a conference operating in mode 2 of route configuration *i*. Suppose there is a change of mode from 2 to 4. If a channel in the reversed path between the old and the new source nodes is not available, video freeze occurs. The reroutable assignment allows the conference to be rerouted
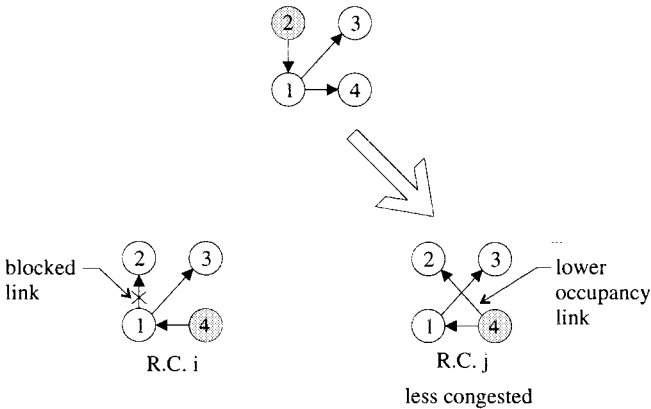
Figure 6. Rerouting conferences.

to a route-configuration that is less congested, say, route-configuration j as shown. In other words, the reroutable assignment helps to distribute the video traffic evenly throughout the network and can reduce the link blocking probability and hence the video freeze probability.

The following is the rerouting procedure:

1. Conferee initiates change of source node.
2. Conference manager notifies control center of the change.
3. Control center computes the least congested route-configuration and reports to the conference manager.
4. Conference manager takes down connections no longer needed and establishes new connections.
5. Conference resumes on the new route-configuration.

We focus our study on the following two reroutable assignment schemes:

NORMAL SCHEME. *A conference is rerouted to the least congested route-configuration whenever there is a change of source node.*

STICKY SCHEME. *A conference stays on its present route-configuration until congestion occurs. When that happens, it is rerouted to the least congested route-configuration. It will stay there until the next congestion occurs and hence the name sticky scheme. This scheme resembles the sticky routing algorithm used in the British Telecom's network.*

Rerouting requires the searching of the least congested route-configurations and the taking down and the establishing of connections. In comparing different reroutable assignment schemes, the rerouting probability, which is equal to the probability that a conference is rerouted when there is a change of source node, can be used to measure the processing overhead required.

4.   EXTENDED CAPACITY AND REROUTING PROBABILITY

The extended capacity is the set of all combinations of conference traffic the network can handle without violating a present QoS requirement when some of the conferences are reroutable. This is in contrast to the *basic* capacity attained when all the conferences are routed by the basic assignment.[10] In the following, we call conferences that are reroutable the reroutable conferences, those not reroutable are called basic conferences.

For basic conferences, define a conference realization r to be a conference of a particular type on a particular route-configuration. Let $\mathbf{R} = \{r^{(1)}, r^{(2)}, \ldots, r^{(Y)}\}$ be the set of all configuration realizations, Y in total. For conference realization $r^{(i)}$, we define the following variables:

$m^{(i)}$: the total number of conference nodes.
$1/x_j^{(i)}$: the mean active period of mode j, $j = 1, 2, \ldots, m^{(i)}$.
$\gamma_{j,h}^{(i)}$: the transition probability from mode j to h, where $j, h = 1, 2, \ldots, m^{(i)}$.
$K^{(i)}$: the number of existing basic conferences, or realization $r^{(i)}$; $\mathbf{K} = (K^{(1)}, K^{(2)}, \ldots, K^{(Y)})$.
$k_j^{(i)}$: the number of basic conferences of realization $r^{(i)}$ on route-configuration j, $j = 1, 2, \ldots, m^{(i)}$; $\mathbf{k}^{(i)} = (k_1^{(i)}, k_2^{(i)}, \ldots, k_{m^{(i)}}^{(i)})$, and $\mathbf{k} = (\mathbf{k}^{(1)}, \mathbf{k}^{(2)}, \ldots, \mathbf{k}^{(Y)})$.

While the route-configurations of reroutable conferences may change, their conference types remain the same. For a particular conference type, we shall call a conference operating in a particular mode under a particular route-configuration a *route-configuration mode*, or RCM for short. Parallel to the case for basic conferences, we define a corresponding set of variables for reroutable conferences by marking them with "$\hat{}$" as follows:

$\hat{m}^{(i)}$: the number of distinct RCMs for type i conferences.
$\hat{K}^{(i)}$: the number of existing type i conferences; $\hat{\mathbf{K}} = (\hat{K}^{(1)}, \hat{K}^{(2)}, \ldots, \hat{K}^{(Z)})$.
$\hat{k}_j^{(i)}$: the number of type i conferences $\mathbf{n}^{(i)}$ operating in RCM j, $j = 1, 2, \ldots, \hat{m}^{(i)}$; $\hat{\mathbf{k}}^{(i)} = (\hat{k}_1^{(i)}, \hat{k}_2^{(i)}, \ldots, \hat{k}_{\hat{m}^{(i)}}^{(i)})$, and $\hat{\mathbf{k}} = (\hat{\mathbf{k}}^{(1)}, \hat{\mathbf{k}}^{(2)}, \ldots, \hat{\mathbf{k}}^{(Z)})$.

With both basic and reroutable assignments allowed, $(\mathbf{K}, \hat{\mathbf{K}})$ is the traffic loaded onto the network and is called a traffic combination. We derive the video freeze probability of the network when it is loaded by a traffic combination $(\mathbf{K}, \hat{\mathbf{K}})$.

4.1.   Limiting Probability Distribution

As the change of modes of basic conferences is governed by an external process independent of the link loadings, the limiting probability of $\mathbf{k}$ is obtained as if there are only basic conferences in the network. Suppose there are only basic conferences. Consider the $K^{(i)}$ conferences of realization $r^{(i)}$ as a subsystem. Let us classify all conferences there by their modes and represent the

conferences in each model by an infinite server queue.‡ The subsystem therefore has a total of $m^{(i)}$ queues for the $m^{(i)}$ modes. Since $k_j^{(i)}$ is the number of conferences of realization $r^{(i)}$ that is operating in mode j, $\mathbf{k}^{(i)} = (k_1^{(i)}, k_2^{(i)}, \ldots, k_{m^{(i)}}^{(i)})$ is the state for this subsystem. The state space $\mathscr{A}^{(i)}$ of this subsystem is given by

$$\mathscr{A}^{(i)} = \left\{ \mathbf{k}^{(i)} \colon \sum_{j=1}^{m^{(i)}} k_j^{(i)} = K^{(i)}, \; k_j^{(i)} \geq 0 \; \forall j \right\} \tag{3}$$

This is a closed Markovian queueing network and its limiting state probability is given in Ref. 14 as

$$p[\mathbf{k}^{(i)}] = \frac{\prod_{j=1}^{m^{(i)}} \left(y_j^{(i)} \right)^{k_j^{(i)}} / \left(k_j^{(h)} \right)!}{\sum_{\mathbf{k}^{(h)} \in \mathscr{A}^{(h)}} \prod_{j=1}^{m^{(h)}} \left(y_j^{(h)} \right)^{k_j^{(h)}} / \left(k_j^{(h)} \right)!} \tag{4}$$

where the $y_j^{(i)}$'s are the solutions of

$$x_j^{(i)} y_j^{(i)} = \sum_{h=}^{m^{(i)}} x_h^{(i)} y_h^{(i)} \gamma_{h,j}^{(i)} \quad j = 1, 2, \ldots, m^{(i)}$$

$$x_1^{(i)} y_1^{(i)} = 1 \tag{5}$$

To generalize the analysis for all conference realizations, we let $\mathbf{k} = (\mathbf{k}^{(1)}, \mathbf{k}^{(2)}, \ldots, \mathbf{k}^{(Y)})$ be the system state. Since conferences are not allowed to change their route-configurations in basic assignments, conferences in different realizations are independent, the limiting probability $p[k]$ is just the product of those for the individual subsystems, i.e.,

$$p[\mathbf{k}] = \prod_{i=1}^{Y} p[\mathbf{k}^{(i)}] \tag{6}$$

We now consider the reroutable conferences. The change of RCMs for reroutable conferences depends on the current link loadings, or the system state. Given a particular $\mathbf{k}$, the activities of the reroutable conferences can be modeled by a continuous time Markov chain with state $\hat{\mathbf{k}}$ and state space $\hat{\mathscr{A}}$ given by

$$\hat{\mathscr{A}} = \left\{ \hat{\mathbf{k}} \colon \sum_{j=1}^{\hat{m}^{(i)}} \hat{k}_j^{(i)} = \hat{K}^{(i)} \text{ for } i = 1, 2, \ldots, Z \right\} \tag{7}$$

‡ The queue length here represents the number of conferences of a particular mode and not the number of conferences actually carried on the links. In other words, it is not restricted by the link capacities as some of the conferences may be experiencing video freeze.

Let $\delta_j^{(i)}(l)$ and $\hat{\delta}_j^{(i)}(l)$ be two indicator functions defined as

$$\delta_j^{(i)}(l) = \begin{cases} 1 & \text{if a basic conference of realization } r^{(i)} \text{ operating in mode } j \\ & \text{requires a channel on link } l, \\ 0 & \text{otherwise}. \end{cases} \tag{8}$$

and

$$\hat{\delta}_j^{(i)}(l) = \begin{cases} 1 & \text{if a reroutable conference of type } i \text{ operating in} \\ & \text{RCM } j \text{ requires a channel on link } l \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Then $\tau(l|(\mathbf{k}, \hat{\mathbf{k}}))$, the number of conferences assigned to link l at state $(\mathbf{k}, \hat{\mathbf{k}})$ is equal to

$$\tau\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big) = \sum_{i=1}^{Y} \sum_{j=1}^{m^{(i)}} k_j^{(i)} \delta_j^{(i)}(l) + \sum_{i=1}^{Z} \sum_{j=1}^{\hat{m}^{(i)}} \hat{k}_j^{(i)} \hat{\delta}_j^{(i)}(l) \tag{10}$$

We measure the congestion of a route-configuration by the loading levels of all the links involved. Let $\hat{\mathscr{L}}_j^{(i)}$ be the set of directed links associated with RCM j of type i conferences. Let $c(l)$ denote the capacity of link l. Let $(x)^+$ be defined as $\max[x, 0]$. Then,

$$d_u^{(i)}\big((\mathbf{k}, \hat{\mathbf{k}})\big) = \min_{l \in \hat{\mathscr{L}}_u^{(i)}} \left[\big(c(l) - \tau\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big)\big)^+ \right] \tag{11}$$

is the minimum number of remaining channels among all the links of RCM u. The larger the value, the less congested is the RCM.

Let $\hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}}))$ be the transition probability from RCM u to v for type i conference at state $(\mathbf{k}, \hat{\mathbf{k}})$. Suppose RCMs u and v correspond to source nodes s_1 and s_2, respectively. The transition probability from mode s_1 to mode s_2 is given in Section 2.1 as γ_{s_1, s_2}. Different reroutable schemes give different sets of $\hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}}))$. For the normal scheme, the relationship between $\hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}}))$ and γ_{s_1, s_2} is given by

$$\hat{\gamma}_{u,v}^{(i)}\big((\mathbf{k}, \hat{\mathbf{k}})\big) = \begin{cases} \dfrac{1}{\eta} \gamma_{s_1, s_2} & \text{if } \eta \text{ RCMs in } \chi_{s_2}^{(i)} \text{ have the same} \\ & \text{maximum } d_v^{(i)}\big((\mathbf{k}, \hat{\mathbf{k}})\big) \text{ value} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\chi_s^{(i)}$ is the set of all RCMs of type i conferences with node s as the source node.

For the sticky scheme, let u^* be the RCM with source node s_2 having the same route-configuration as the current RCM u. A conference is only rerouted

if RCM u^* is congested, or $d_{u^*}^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) = 0$. Therefore, we have

$$
\hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) = \begin{cases} \gamma_{s_1, s_2} & \text{if } v = u^* \text{ and } d_{u^*}^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) > 0 \\[2mm] \dfrac{1}{\eta} \gamma_{s_1, s_2} & \text{if } d_{u^*}^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) = 0 \text{ and } \eta \text{ RCMs in } \chi_{s_2}^{(i)} \text{ have the same} \\[2mm] & \quad \text{maximum } d_v^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) \text{ value} \\[2mm] 0 & \text{otherwise} \end{cases}
$$

(13)

Given a particular $\mathbf{k}$, let $q_{\hat{\mathbf{k}}, \hat{\mathbf{k}}'|\mathbf{k}}$ be the rate for the system to change from state $\hat{\mathbf{k}}$ to state $\hat{\mathbf{k}}'$. Then,

$$
q_{\hat{\mathbf{k}}, \hat{\mathbf{k}}'|\mathbf{k}} = \begin{cases} \hat{k}_u^{(i)} \cdot \hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}})) \cdot \mu & \text{if } \hat{k}_u^{(i)\prime} = \hat{k}_u^{(i)} - 1, \hat{k}_v^{(i)\prime} = \hat{k}_v^{(i)} + 1 \\ & \quad \hat{k}_w^{(i)\prime} = \hat{k}_w^{(i)} \; \forall w \neq u, v \\ & \quad \hat{\mathbf{k}}^{(h)\prime} = \hat{\mathbf{k}}^{(h)} \; \forall h \neq i \\ -\sum_{\forall \hat{\mathbf{k}}'' \neq \hat{\mathbf{k}}} q_{\hat{\mathbf{k}}, \hat{\mathbf{k}}''|\mathbf{k}} & \text{if } \hat{\mathbf{k}}' = \hat{\mathbf{k}} \\ 0 & \text{otherwise} \end{cases}
$$

(14)

Define the transition rate matrix as $Q_{\mathbf{k}} = [q_{\hat{\mathbf{k}}, \hat{\mathbf{k}}'|\mathbf{k}}]$. Let $p[\hat{\mathbf{k}}|\mathbf{k}]$ be the limiting probability for state $\hat{\mathbf{k}}$ given $\mathbf{k}$, and $\pi_k = (p[\hat{\mathbf{k}}_1|\mathbf{k}], \ldots, p[\hat{\mathbf{k}}_{|\hat{\mathscr{A}}|}|\mathbf{k}])$. Then, $p[\hat{\mathbf{k}}|\mathbf{k}]$ can be obtained by solving the following set of equations,

$$
\pi_k Q_k = 0 \tag{15}
$$

$$
\sum_{\forall \hat{\mathbf{k}}} p[\hat{\mathbf{k}}|\mathbf{k}] = 1 \tag{16}
$$

Removing the conditioning on $\mathbf{k}$, we obtain

$$
p[(\mathbf{k}, \hat{\mathbf{k}})] = p[\hat{\mathbf{k}}|\mathbf{k}] \, p[\mathbf{k}] \tag{17}
$$

It is easy to see that the size of Q grows very fast with the number of nodes N. For example, if we assume that the number of route-configurations to those that have at most three hops between nodes and that there are k conferences of each type in the system, the state space is of size

$$
\prod_{j=2}^{N} \binom{k + j(1 + (j-2)(j-3)) - 1}{k}^{\binom{N}{j}} \tag{18}
$$

Fortunately the matrix is sparse. However, even then, for N larger than 4 or 5, efficient approximate techniques need to be found. This however is beyond the scope of this paper.

4.2. Video Freeze Probability

Given the system is in state $(\mathbf{k}, \hat{\mathbf{k}})$, the blocking probability of link l is given by

$$b\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big) = \begin{cases} \dfrac{\tau\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big) - c(l)}{\tau\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big)} & \text{if } \tau\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big) > c(l) \\ 0 & \text{otherwise} \end{cases} \qquad (19)$$

where $\tau(l|(\mathbf{k}, \hat{\mathbf{k}}))$ is the number of conferences assigned to link l at state $(\mathbf{k}, \hat{\mathbf{k}})$ and is given by (10). Let $v_j^{(i)}((\mathbf{K}, \hat{\mathbf{K}}))$ be the video freeze probability for the basic conferences operating in mode j of realization $r^{(i)}$ under traffic combination $(\mathbf{K}, \hat{\mathbf{K}})$. Let $\mathscr{L}_j^{(i)}$ be the set of directed links associated with mode j of realization $r^{(i)}$. Given that the system is in state $(\mathbf{k}, \hat{\mathbf{k}})$, such a conference experiences video freeze if any link in $\mathscr{L}_j^{(i)}$ is blocked, the probability of which is equal to $(1 - \prod_{l \in \mathscr{L}_j^{(i)}}(1 - b(l|(\mathbf{k}, \hat{\mathbf{k}}))))$. Removing the conditioning on $(\mathbf{k}, \hat{\mathbf{k}})$ yields

$$v_j^{(i)}\big((\mathbf{K}, \hat{\mathbf{K}})\big) = \sum_{\forall(\mathbf{k}, \hat{\mathbf{k}})} \left\{ 1 - \prod_{l \in \mathscr{L}_j^{(i)}} \Big(1 - b\big(l|(\mathbf{k}, \hat{\mathbf{k}})\big)\Big) \right\} p\big[(\mathbf{k}, \hat{\mathbf{k}})\big] \qquad (20)$$

However, as there is a one-to-one correspondence between a mode for a particular conference realization and a RCM for a particular conference type, $v_j^{(i)}((\mathbf{K}, \hat{\mathbf{K}}))$ is also the video freeze probability for reroutable conferences operating in a particular RCM with a particular conferee distribution. Therefore it is necessary to define another set of variables for the video freeze probability of reroutable conferences.

4.3. Extended Capacity Space

The extended capacity space Ω is defined as

$$\Omega = \Big\{ (\mathbf{K}, \hat{\mathbf{K}}) \colon v_j^{(i)}\big((\mathbf{K}, \hat{\mathbf{K}})\big) \le v' \text{ for } i = 1, 2, \ldots, Y; j = 1, 2, \ldots, m^{(i)} \Big\} \quad (21)$$

where v^* is a given QoS requirement. This means that for all $(\mathbf{K}, \hat{\mathbf{K}})$ in Ω, the video freeze probability is at most v^* for all conferences.

Call admission can be performed in two ways: 1. the video freeze probabilities under a specific traffic combination are computed in real time to see if they are all smaller than the present QoS requirement, if so, the new conference call is admitted; 2. the capacity space Ω is precomputed and stored in the control center. A check if the traffic combination (with the new conference added) falls within Ω or not determines if the new conference is admissible or not. It is easy to see that approach 1 is feasible only for very small networks and approach 2 would not be feasible for N large. However, for large networks, it should be fairly straightforward to ask the network to "learn" its own capacity. In other words, we can keep adding different combinations of conferences until the QoS

is barely satisfied and record those values. This can be done either by computer simulation or on a real network.

4.4.   Rerouting Probability

Let $P_R[(\mathbf{k}, \hat{\mathbf{k}})]$ be the probability of rerouting given the system is at state $(\mathbf{k}, \hat{\mathbf{k}})$. This is the probability that a conference is rerouted to another route-configuration when there is a change of source node. Let $RC(u)$ denote the route-configuration of RCM u. At state $(\mathbf{k}, \hat{\mathbf{k}})$, there are $\hat{k}_u^{(i)}$ reroutable conferences of type i operating in RCM u. The transition probability from RCM u to RCM v is $\hat{\gamma}_{u,v}^{(i)}((\mathbf{k}, \hat{\mathbf{k}}))$. Therefore we have

$$P_R\big[(\mathbf{k}, \hat{\mathbf{k}})\big] = \sum_{i=1}^{Z} \left(\sum_{(u,v) \in \Psi} \hat{\gamma}_{u,v}^{(i)}\big((\mathbf{k}, \hat{\mathbf{k}})\big) \cdot \frac{\hat{k}_u^{(i)}}{\hat{K}_{\text{sum}}} \right) \qquad (22)$$

where $\Psi = \{(u, v) | RC(u) \neq RC(v)\}$ and $\hat{K}_{\text{sum}} = \sum_{i=1}^{Z} \hat{K}^{(i)}$ is the sum of all reroutable conferences loaded onto the network. Unconditioning on $(\mathbf{k}, \hat{\mathbf{k}})$ yields

$$P_R = \sum_{\forall(\mathbf{k}, \hat{\mathbf{k}})} \left(\sum_{i=1}^{Z} \left(\sum_{(u,v) \in \Psi} \hat{\gamma}_{u,v}^{(i)}\big((\mathbf{k}, \hat{\mathbf{k}})\big) \cdot \frac{\hat{k}_u^{(i)}}{\hat{K}_{\text{sum}}} \right) \right) \cdot p\big[(\mathbf{k}, \hat{\mathbf{k}})\big] \qquad (23)$$

4.5.   Example

Let us consider a specific conference type in a 3-node network having exactly one conferee at each node. For simplicity, we consider only two route-configurations. Since there are three modes of operation for each route config-uration, there are a total of six RCMs shown in Figure 7 for reroutable conferences. Let each directed link have a capacity of five channels and let v^*
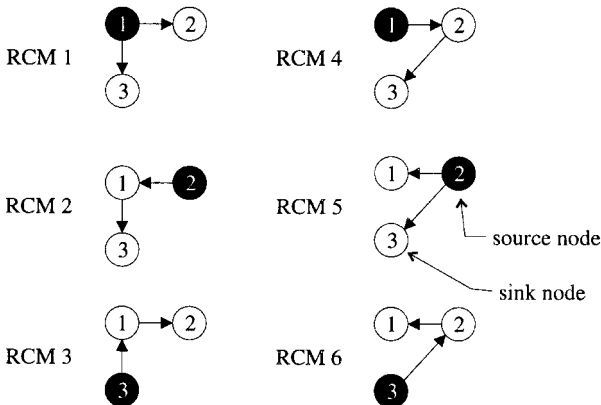

Figure 7.   The six RCMs of the example.

Table II. Basic capacity space for the example.

	$K^{(1)}$					
	0	1	2	3	4	5
max $K^{(2)}$	5	5	4	3	2	1

be set to 0.01. Suppose there are only basic conferences in the network, the basic capacity space is a set of two-dimensional vectors $(K^{(1)}, K^{(2)})$. This is shown in Table II as the maximum value of $K^{(2)}$ for various values of $K^{(1)}$. The maximum number of conferees the network can accommodate is counted to be six.

Next, suppose there are only reroutable conferences in the network. Let $\hat{K}$ be the number of reroutable conferences loaded to the network; v_i, $(i = 1, 2, \ldots, 6)$ be the video freeze probability for RCM i. Table III shows the video freeze probabilities and rerouting probabilities for $\hat{K}$ ranges from 6 to 8. We see that the maximum number of conferences that can be accommodated is 7 for both the normal and the sticky schemes. Thus in this example, the extended capacity is one unit larger than the basic capacity. Considering the cases of $\hat{K} = 6$ and 7, the normal scheme gives slightly smaller video freeze probabilities than the sticky scheme. The rerouting probabilities for the normal scheme, however, are significantly larger. When $\hat{K} = 8$, the network is already heavily loaded and both schemes give similar performance.

5. BLOCKING PROBABILITY

In the last section, we derived the video freeze probability of a conferencing network under a particular traffic combination $(\mathbf{K}, \hat{\mathbf{K}})$. The network under consideration is a closed system with no conference arrival and departure. In this section, we consider both the arrival and departure of conferences to the

Table III. Video freeze probability and rerouting probability for the normal and the sticky schemes.

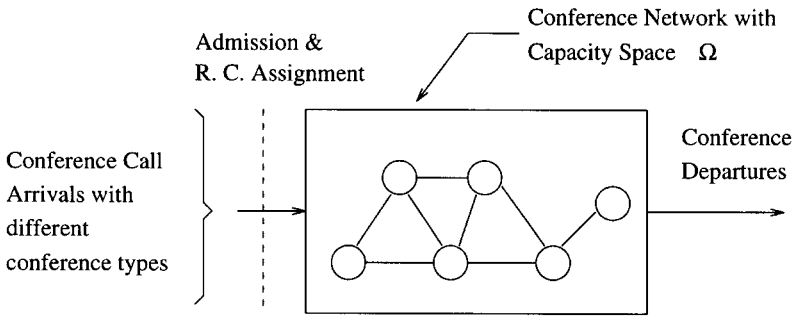|	Normal Scheme						
$\hat{K}$	v_1	v_2	v_3	v_4	v_5	v_6	P_R
6	0.0008	0.0008	0.0006	0.0008	0.0008	0.0006	0.0940
7	0.0041	0.0042	0.0032	0.0042	0.0041	0.0032	0.1717
8	0.0137	0.0143	0.0115	0.0143	0.0137	0.0115	0.01083
	Sticky Scheme						
$\hat{K}$	v_1	v_2	v_3	v_4	v_5	v_6	P_R
6	0.0010	0.0011	0.0009	0.0011	0.0010	0.0009	0.0103
7	0.0049	0.0052	0.0041	0.0052	0.0049	0.0041	0.0396
8	0.0137	0.0143	0.0115	0.0143	0.0137	0.0115	0.0909

Figure 8.   Arrival and departure of conferences.

network. Figure 8 shows the system model. The box represents the conferencing network. When there is a new conferenced arrival, the call admission scheme checks whether the new call can be admitted or not. The admission decision is based on the new conference type, the capacity space Ω and the current traffic combination. If the new conference can be assigned to some route-configurations such that the resulting traffic combination is inside Ω, it is admitted. Otherwise, it is blocked and lost. We now derive the call blocking probability for conferences of a specific type.

   We assume the arrival of conference calls is a Poisson process. Let $\lambda^{(i)}$ be the arrival rate of type i conferences, $(i = 1, 2, \ldots, Z)$. Speech duration of any speaker is assumed to be exponentially distributed with mean μ^{-1}. When a conferee finishes speaking, the conference ends immediately with probability p and continues with probability $1 - p$. If the conference continues, the new speaker is equally likely to be any one of the other conferees. Under this assumption, the duration of a conference call in the network is a geometric sum of independent and identically distributed exponential random variables. Its distribution is exponential with mean $(\mu p)^{-1}$ as discussed in Section 2.1.

   We derive the call blocking probability for a network in which all conferences are reroutable. The conference call blocking probabilities for a network with both basic and reroutable conferences can be derived in a similar fashion. The system can be modeled by a continuous time Markov chain. There are altogether Z distinct conference types, a traffic combination is defined by $\mathbf{K} = (K^{(1)}, K^{(2)}, \ldots, K^{(Z)})$, where $K^{(i)}$ is the number of type i conferences loaded onto the network. Let $\mathbf{h} = (h^{(1)}, h^{(2)}, \ldots, h^{(Z)})$ be the system state with $h^{(i)}$ equal to the number of existing type i conferences. Define $\mathbf{e}^{(i)}$, $(i = 1, 2, \ldots, Z)$, to be a unit vector of size Z with the ith component equal to 1 and all other components equal to 0.

   Define the indicator function $I_{\mathbf{h}}$ as

$$I_{\mathbf{h}} = \begin{cases} 1 & \text{if } \mathbf{h} \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

To construct the transition rate matrix $\mathbf{Q} = [q_{\mathbf{h},\mathbf{h}'}]$ where $q_{\mathbf{h},\mathbf{h}'}$ is the transition state from state $\mathbf{h}$ to $\mathbf{h}'$, consider the change of state from $\mathbf{h}$ to $\mathbf{h} + \mathbf{e}^{(i)}$. This change corresponds to the arrival of a type i conference. Therefore,

$$q_{\mathbf{h},\mathbf{h}+\mathbf{e}^{(i)}} = \lambda^{(i)} \cdot I_{\mathbf{h}+\mathbf{e}^{(i)}}$$

The change of state from $\mathbf{h}$ to $\mathbf{h} - \mathbf{e}^{(i)}$ corresponds to the termination of a type i conference. So we have

$$q_{\mathbf{h},\mathbf{h}-\mathbf{e}^{(i)}} = h^{(i)} \cdot \mu p$$

With that, the limiting probability $p[\mathbf{h}]$ can be solved as usual. To find the blocking probability, define $\Theta^{(i)}$ as the *capacity boundary* for type i conferences, or

$$\Theta^{(i)} = \{\mathbf{h}: \mathbf{h} \in \Omega, \mathbf{h} + \mathbf{e}^{(i)} \notin \Omega\} \tag{25}$$

When the system is in state $\mathbf{h} \in \Theta^{(i)}$, it is not able to admit any type i conference. In other words, a new type i conference arrival is blocked if the system is on the capacity boundary $\Theta^{(i)}$.

The blocking probability $P_B(i)$ for type i conferences can be obtained by summing up the limiting probabilities of all the states in $\Theta^{(i)}$, or

$$P_B(i) = \sum_{\forall \mathbf{h} \in \Theta^{(i)}} p[\mathbf{h}] \tag{26}$$

For the previous example, as there is only one conference type, the model reduces to a one-dimensional Markov chain. Let h be the number of conferees in the network and C be the maximum number of conferences that can be loaded to the network without violating QoS requirement. The limiting probability $p[h]$ can be shown to be[14]

$$p[h] = \frac{1}{h!}\left(\frac{\lambda}{\mu p}\right)^h \cdot p[0] \qquad h = 0, \ldots, C \tag{27}$$

where $p[0]$ is given by

$$p[0] = \left[\sum_{h=0}^{C} \frac{1}{h!}\left(\frac{\lambda}{\mu p}\right)^h\right]^{-1} \tag{28}$$

The conference call blocking probability is $P_B = p[C] = p^7$ as given in (27).

6.   CONCLUSION

This paper presents an analysis of multipoint video conferencing in a communication network. A conferencing network model is constructed which includes a formal classification of conference traffic. Conferences are character-

ized by types and each type can be realized by different route-configurations. A conference on a particular route-configuration has different modes of operation.

Two conferencing network management functions, namely, admission and route-configuration assignment, are described. The basic and reroutable route-configuration assignments are introduced. The basic assignment simply says that the route-configuration of a conference, once assigned, remains the same throughout the conference session. Two reroutable assignment schemes are studied. With the normal scheme, a conference is rerouted to the least congested route-configuration whenever there is a change of source node. With the sticky scheme, a conference stays on its present route-configuration until congestion occurs. When this happens, it is rerouted to the least congested route-configuration. However, as a reroutable assignment can distribute traffic more evenly onto the network, the link blocking probability can be reduced as compared to basic assignment. This results in a smaller video freeze probability and a larger capacity space. We derived the video freeze probability, the rerouting probability and the extended capacity space. The normal scheme and the sticky scheme given comparable video freeze probabilities. However, the sticky scheme is superior because it gives significantly smaller rerouting probability.

The computational complexity for the extended capacity is very high. Therefore for networks larger than five nodes say, efficient approximate analysis is needed.

References

1. M. Hopkins, G. Louth, H. Bailey, R. Yellon, A. Ajibulu, and M. Niva, "A multi-faceted approach to forecasting broadband demand and traffic," *IEEE Comm. Mag.*, 36–43 (Feb. 1995).
2. K. Stordahl and E. Murthy, "Forecasting long-term demand for services in the residential market," *IEEE Comm. Mag.*, 44–49 (Feb. 1995).
3. S.R. Ahuja and J.R. Ensor, "Coordination and control of multimedia conferencing," *IEEE Comm. Mag.*, **30**(5), 38–43 (May 1992).
4. W.J. Clark, "Multipoint multimedia conferencing," *IEEE Comm. Mag.*, **30**(5), 44–50 (May 1992).
5. T. Ohmori, K. Maeno, S. Sakata, H. Fukuoka, and K. Watabe, "Cooperative control for sharing applications based on distributed multiparty desktop conferencing system: MERMAID," *IEEE ICC'92*, 1992, pp. 1069–1075.
6. S. Wilbur, B. Hewitt, and S. Ing, "A multimedia system for flexible coorpation," *IEEE ICC'92*, 1992, pp. 1087–1091.
7. S. Sabri and B. Prasada, "Video conferencing systems," *Proc. IEEE*, **73**(4), 671–688 (Apr. 1985).
8. G. Romaha, "System aspects of multipoint videoconferencing," *Proc. IIII GLOBE-COM*, 1987, pp. 723–725.
9. M.J. Ferguson and L. Mason, "Network design for a large class of teleconferencing systems," *IEEE Trans. Commun.*, **32**(7), 789–796 (July 1984).
10. T.K. Chan, "Traffic engineering for multipoint videoconferencing," M.Ph. Thesis, CUHK, 1995.
11. T.S. Yum, M.S. Chen, and Y.W. Leung, "Video bandwidth allocation for multimedia teleconferences," *IEEE Trans. Commun.*, 457–465 (Feb. 1995).

12. P. Sen, B. Maglaris, N.E. Rikli, and D. Anastassiou, "Models for packet switching of variable-bit-rate video sources," *IEEE J. Select. Areas Commun.*, **7**(5) (June 1989).

13. W.C. Lau, A. Erramilli, J.L. Wang, and W. Willinger, "Self-similar traffic parameter estimation: A semiparametric periodogram-based algorithm," *Proc. IEEE Globecom*, 1995, pp. 2225−2231.

14. L. Kleinrock, *Queuing Systems*, *Vol*. 1*: Theory*, Wiley Interscience, 1975, p. 152.

15. R.J. Wilson and J.J. Watkins, *Graphs: An Introductory Approach*, Wiley, 1990, pp. 190−193.