Video Bandwidth Allocation for Multimedia Teleconferences

Tak-Shing Yum, Mon-Song Chen and Yiu-Wing Leung

Abstract— **To ensure the quality of a multimedia teleconference, it is essential that sufficient bandwidth be allocated for its use. In this paper a conference traffic model is formulated and link level and conference level congestion measures are derived. Motivated by the advantages of sharing transmission resources in TASI related voice communication systems, an analogous *transmission policy* for conference videos is proposed. The quantification of conference traffic also enables us to set an *admission policy* so that the network can accommodate as many conferences as possible without violating conference quality constraints.**

I. Introduction

A multimedia teleconference is an information exchange system for people to conduct meetings effectively without leaving their offices. Sitting in front of their workstations, conferees can see each other via real-time motion videos on their multimedia workstation displays, talk and listen to all the conversation via real-time audio, and watch presentations via an on-line Electronic Blackboard.

The effectiveness of a multimedia teleconference should be evaluated based on how much it can emulate a face-to-face meeting. Audio in an multimedia teleconference should be unconstrained so that everybody can talk to and be heard by everybody else. This is possible because audio is a bandwidth efficient medium for people to exchange information and most conferees exercise constraints as to when to speak and when not to. Video, on the other hand, is quite different because of its large transmission bandwidth requirement. Depending upon variables such as dimension, format, and compression, transmission of a good quality compressed video requires at least 1.5 Mbits/sec. The bit rate would be one to two orders of magnitude larger if video is not compressed. In packet switching environment, good video quality means small video packet loss and delay. This calls for careful bandwidth allocation to prevent the network from overload. Two separate but relevant issues need to be addressed in video bandwidth allocation: one is the *admission policy*, which decides whether or not to admit a new conference, and the other is the *transmission policy*, which controls the video transmissions of existing conferences.

Paper approved by J. W. Wong, the Editor for Wide Area Networks of the IEEE Communications Society. Manuscript received: October 30, 1991; revised April 7, 1993. This paper was presented in part at the IEEE ICC '91, Denver, Colorado.

T. S. Yum is with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong.

M. S. Chen is with IBM T.J.Watson Research Center, P.O.Box 704, Yorktown Heights, NY 10598.

Y. W. Leung is with the Department of Computing, Hong Kong Polytechnic, Hong Kong.

IEEE Log Number 9410077.

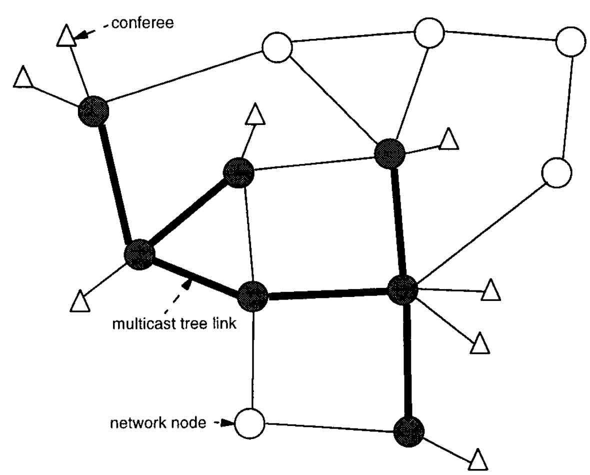

Fig. 1. The topological layout of a conference in a network.

Bandwidth allocation for conferencing is unique due to its multicasting nature. In a typical conference there could be multiple conference members homing in a particular conference node. The set of nodes involved in a conference could be linked by a minimum multicast tree as shown in Fig. 1. It is reasonable to assume that one video transmission for one conference is sufficient in most practical situations. The video would most likely be used to show the current "speaker". This is obviously less than ideal but would serve the conferencing purpose with minimum bandwidth allocation from the network. With this assumption the bandwidth utilization would be only 50% on the average if each directed link in the multicast tree is allocated with the full bandwidth of a video transmission. The bandwidth utilization can be improved by exploring the benefit of statistical multiplexing, as has been done in TASI [1] related voice communication systems.

In this paper, we propose a bandwidth allocation architecture which includes both admission and transmission policies. A conference traffic model is proposed and solved for determining the control parameters used in these policies.

II. Bandwidth Allocation Architecture

In this section we propose a bandwidth allocation architecture for a network to regulate conference traffic. The

0090-6778/95$4.00

architecture is simple and suitable for distributed implementations. We will first describe the two necessary assumptions and then the architecture.

The first assumption is that each conference has exactly one video multicast transmission at all time. The source of the video multicast changes from one conferee to the other according to an external procedure which is completely independent of other conferences and the status of the network. Additional video communications, such as private point-to-point or a smaller scale multicast transmissions, are allowed whenever possible but not guaranteed, and should have lower priority than the guaranteed video multicast. These additional traffic, therefore, will not be included in our conference traffic model.

The second assumption is that all videos are transmitted in the same format, and 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 setup as in circuit switched environment, or can be done via known queueing techniques [2-3] if video are packetized and transmitted in store-and-forward fashion as in packet switching environment. The queueing analysis in the latter case needs as inputs the calibrated video quality in terms of packet 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.

There are two aspects in the bandwidth allocation architecture: the transmission policy, which controls video transmission of admitted conferences, and the admission policy, which decides whether or not to admit a new conference:

A. Transmission Policy

Before starting a video transmission, the source of the video must make a request to and obtain the permission from every directed network link included in the multicast tree. A network link, upon receiving a request for starting a new video transmission over it, checks to see if there is sufficient slack bandwidth to accommodate the new additional video traffic, and responds accordingly. The source can start sending video data only if every link responds positively; otherwise, the source is prohibited from sending any video data into the network. The consequence of the prohibition is a temporary video-freeze for the affected conference. As soon as all links needed for the video transmission become available, the video transmission proceeds. An alternative is to try to accommodate as many video transmissions as possible by asking each video source to reduce its data rate. This rate-reduction approach is interesting in that service degradation is shared indiscreetly by all conferences sharing the congested link. The disadvantage is that video quality may fluctuate when the number of video transmissions exceed a certain threshold and that more complicated multirate video coding is required. In comparison, we argue that it may be more acceptable for conferees to experience a momentary video-freeze at the beginning of a video session than to tolerate unsteady video quality. This video-freeze approach is conceptually similar to the clipping of voice in TASI systems except that it is less annoying as it can occur only at the beginning of a video session, not after every pause. Guaranteeing acceptable mean video-freeze duration and probability of occurrence lies on the enforcement of the Admission Policy.

B. Admission Policy

When a request for a new conference is received, every link in the specified connectivity must be informed. Each link calculates the video-freeze statistics if the new conference were to be admitted. Specifics of how to calculate video-freeze statistics will be discussed in detail in Section IV. If the results indicate that some constraints will be violated, the link responds negatively. Otherwise, the link responds with the new video-freeze statistics. If there is a negative response from any link, the conference is rejected. If all responses are positive, then a network wide analysis is performed to estimate the performance to be experienced by the prospective new conference. The network wide analysis will be discussed in Section V. The new conference is admitted if the performance is acceptable.

III. Conference Traffic Model

A conference on a network is characterized by three attributes: the set of nodes with conferees attached, the activity levels of individual conferees, and the connectivity of the nodes involved in the conference. Consider a specific conference with K conferees. Let these K conferees be homed onto a set of nodes, which in turn, are connected by a multicast tree. Consider any two adjacent nodes say node A and node B on the tree. Let link AB be the one way link from node A to node B and link BA be that from node B to node A. Let there be F conferees to the left of link AB (i.e. whose transmissions go through link AB) and $K - F$ conferees to the right (i.e. whose transmissions go over link BA). Without loss of generality let us label the conferees such that conferees 1 to F are to the left of link AB and conferees $F+1$ to K are to the right. Let a_i be the probability that conferee i is the speaker.

A conferee is *active* if it is the speaker of its conference and thus has the priviledge of sending video. It may not, however, be actually sending video because the network may be out of bandwidth temporarily. An active conferee becomes inactive when a different conferee takes over the role of a speaker. As such each conferee cycles through active and inactive periods.

Let the active period of a conferee be exponentially distributed with a mean of $1/\mu$ second. Then the activities of the conference on the network can be modeled by a K state Markov chain with state i denoting the event conferee i is active. By definition a_i is the steady state probability of state i. Let us create two regions such that region L (left) contains state 1 to state F and region R (right) contains state $F+1$ to state K.

When a conferee in region L is active, the video transmission of the conference will go over link AB. The transmission will continue until a speaker in region R becomes active. To determine the amount of traffic contributed by a conference to link AB, we first need to find the statistics of the channel holding time which is the time a conference is active on link AB. Referring to the Markov chain model, the channel holding time is readily seen to be the time the process stays in regions L or the passage time in region L. To determine the passage time statistics we first need to determine the state transition probabilities.

A. Transition Probabilities

Modeling multiparty video conferences is in general fairly difficult. First of all, there are many types of conferences and conferees may behave very differently in each type. Secondly, human beings tend to adapt to the limitations of the electronic media; and so the exact statistical behavior for a specific conferencing system cannot be known until the system has been used by sufficient number of people for a sufficiently long time. In our case, to determine the K^2 unknown transition probabilities in the Markov chain model, we make the simplifying assumption that whom *the next speaker will be* is independent of whom *the present speaker is.* While some conferees do tend to go active after certain speakers but not others, modeling this behavior would increase the dimension of the Markov chain. Lacking any detail statistics for the more detail model, we argue that the above independent transition assumption is reasonable, for estimating the channel holding time statistics of a conference.

Let p_{ij} $(i/neqj)$ be the transition probability from state i to state j. As all conferees go through cycles of active and inactive periods, $p_{ii} = 0$ $(i = 1, 2, \cdots, k)$. The independent transition assumption stipulates that for distinct i, j, m and n,

$$\frac{p_{in}}{p_{in}} = \frac{p_{jm}}{p_{jn}} \tag{1}$$

In other words, p_{im}/p_{in} is independent of i, or the relative rates of transitions to other states is independent of the originating state. This implies that there exists functions $g(i)$ and $h(m)$ such that

$$\begin{aligned} p_{im} &= g(i)h(m) \\ p_{in} &= g(i)h(n) \end{aligned} \tag{2}$$

so that the ratio p_{im}/p_{in} has its dependency on i cancelled. To find $g(i)$ we note that

$$\sum_{m=1,m\neq i}^{K} p_{im} = g(i) \sum_{m=1,m\neq i}^{K} h(m) = 1$$

Solving for $g(i)$, we have

$$g(i) = \frac{1}{H - h(i)}$$

where

$$H = \sum_{m=1}^{K} h(m)$$

Substitute into (2) and divide through by H, we have

$$p_{im} = \frac{r_m}{1 - r_i} \qquad \text{if } i \neq m \tag{3}$$

where $r_m = h(m)/H$. Note that $h(j)$ can be interpreted as the relative rate of transitions to state j and r_j can be interpreted as the unconditioned transition probability to state j. This is in contrast to p_{ij} which is the transition probability to state j conditioned on the current state being i.

To solve for the set of r_m's, we formulate a set of balance equations for the Markov chain by equating the flow, i.e. the product of transition rate and state probability, of entering and leaving a state as follows:

$$\mu a_i = r_i \sum_{j=1,j\neq i}^{K} \frac{a_j}{1 - r_j}\mu \qquad i = 1, 2, \cdots, K \tag{4}$$

Subtracting the equation for a_j from that for a_i in (4), we have

$$\begin{aligned} \frac{a_i}{r_i} - \frac{a_j}{r_j} &= \sum_{k=1,k\neq i}^{K} \frac{a_k}{1 - r_k} - \sum_{k=1,k\neq j}^{K} \frac{a_k}{1 - r_k} \\ &= \frac{a_j}{1 - r_j} - \frac{a_i}{1 - r_i} \end{aligned} \tag{5}$$

Let $U_i = r_i(1 - r_i)$. Then, the r_i's can be solved as

$$r_i = \frac{1 \pm \sqrt{1 - 4U_i}}{2} \qquad i = 1, 2, \cdots, K \tag{6}$$

and (5) can be rewritten as

$$U_j = \frac{a_j}{a_i} U_i \tag{7}$$

It is shown in [7] that if $a_i > a_j$, then $r_i > r_j$. Since all r_i's sum to 1, at most one of the r_i can be larger than 0.5. Denote the r_i corresponding to the largest a_i as r_M. Then

$$r_i = \frac{1 - \sqrt{1 - 4U_i}}{2} \qquad i = 1, 2, \cdots, M-1, M+1, \cdots, K$$

$$r_M = \begin{cases} \frac{1 - \sqrt{1 - 4U_M}}{2} & \text{if } r_M < 0.5 \\ \frac{1 + \sqrt{1 - 4U_M}}{2} & \text{if } r_M \geq 0.5 \end{cases} \tag{8}$$

First assume r_M is smaller than 0.5. Then the normalization equation for r_i's gives

$$\sum_{j=1}^{K} r_j = \sum_{j=1}^{K} \frac{1 - \sqrt{1 - 4U_j}}{2} = 1 \tag{9}$$

Substitute (7) in (9) for all U_j and set $i = 1$, U_1 can be solved numerically. Through (7) all the other U_i's and hence the r_i's (through (8)) are solved. If the calculated $(r_1 + r_2 + \cdots + r_K)$ equals to one, the earlier assumption that $r_M < 0.5$ is valid and we have obtained the solution. If the sume of r_i is smaller than 1, then r_M must be at least 0.5. We then take

$$r_M = \frac{1 + \sqrt{1 - 4U_M}}{2}$$

and repeat the above procedures to obtain the other r_i's.

B. Mean Channel Holding Time on Link AB

Having determined the transition probabilities, we can now solve the passage time. Let $\bar{N}_i$ be the average number of transitions needed to escape from state i to any state in the other region. Then using the method outlined in [4], we can immediately obtain

$$\bar{N}_i = \begin{cases} (\sum_{j=F+1}^{K} \frac{r_j}{1-r_i}) + \frac{1}{1-r_i}\sum_{k=1,k\neq i}^{F} r_k(1+\bar{N}_k) \\ \qquad i = 1,2,\cdots,F \\ (\sum_{j=1}^{F} \frac{r_j}{1-r_i}) + \frac{1}{1-r_i}\sum_{k=F+1,k\neq i}^{K} r_k(1+\bar{N}_k) \\ \qquad i = F+1,\cdots,K \end{cases} \tag{10}$$

These two independent sets of equations can be solved recursively [7] to be:

$$\bar{N}_i = \frac{1-\sum_{j=1}^{F} r_j^2}{\sum_{j=K+1}^{K} r_j} - r_i \quad i = 1,2,\cdots,F \tag{11}$$

$$\bar{N}_i = \frac{1-\sum_{j=F+1}^{K} r_j^2}{\sum_{j=1}^{F} r_j} - r_i \quad i = F+1,\cdots,K \tag{12}$$

To check, let all $r_i = 1/K$ and substitute into (11) we obtain

$$\bar{N}_i = \frac{1}{1-(F-1)/(K-1)} \quad i = 1,2,\cdots,F \tag{13}$$

From an alternate argument, since $(F-1)/(K-1)$ is the probability that the process remains in region L after a transition, the expected number of transitions needed before the process leaves region L given that it starts from state i is

$$\bar{N}_i = \sum_{j=1}^{\infty}(\frac{F-1}{K-1})^{j-1}(1-\frac{F-1}{K-1})$$

This sum converges to give the expression in (13).

As the process starts in state i with probability $a_i/(a_1 + a_2 + \cdots + a_F)$, we can remove the conditioning on i in (11) and multiply (11) by $1/\mu$ to obtain $E[Y]$, the mean channel holding time on link AB as

$$E[Y] = \frac{1}{\mu}(\sum_{i=1}^{F} a_i)^{-1}\sum_{i=1}^{F} a_i\bar{N}_i$$

Similarly, $E[Z]$, the mean time the process stays in region R, or the mean channel holding time for the conference on link BA is

$$E[Z] = \frac{1}{\mu}(\sum_{i=F+1}^{K} a_i)^{-1}\sum_{i=F+1}^{K} a_i\bar{N}_i$$

C. Channel Holding Time Distribution

Given that the process is in state i of region L, the probability of leaving region L in one transition is just the sum of the transition probabilities from state i to all the states in region R, or

$$P[N_i = 1] = \frac{r_{F+1} + r_{F+2} + \cdots + r_K}{1-r_i} \quad i = 1,2,\cdots,F \tag{14}$$

To reach region R in k steps, the process must first stop by one of the states in region L before moving to region R. Therefore

$$P[N_i = k] = \sum_{j=1,j\neq i}^{F} \frac{r_j}{1-r_i}P[N_j = k-1] \quad k = 2,3,\cdots \tag{15}$$

So starting from $P[T_i = 1]$, the entire conditional distribution can be computed recursively. Removing the conditioning on i, we have

$$P[N_{AB} = k] = \frac{\sum_{i=1}^{F} P[N_i = k]a_i}{\sum_{j=1}^{F} a_j} \quad k = 1,2,\cdots$$

Let $X_1, X_2, \cdots$ be exponential i.i.d. random variables representing the successive talking durations of the conferees in region L. Then the channel holding time Y_{AB} on link AB is

$$Y_{AB} = X_1 + X_2 + \cdots + X_{N_{AB}}$$

Its density function $f_{Y_{AB}}$ is derived in [5] as

$$f_{Y_{AB}}(t) = \mu e^{-\mu t}\sum_{k=1}^{\infty}\frac{(\mu t)^{k-1}}{(k-1)!}P[N_{AB} = k] \quad t > 0 \tag{16}$$

Consider a special case when $a_1 = a_2 = \cdots = a_F$ and $a_F + 1 = a_F + 2 = \cdots = a_K$. This implies that $r_1 = r_2 = \cdots = r_F$ and $r_F + 1 = r_F + 2 = \cdots = r_K$. Substitute into (15) and solving the difference equation, $P[N_i = k]$ is obtained as:

$$P[N_i = k] = \begin{cases} (1-\alpha)^{k-1}\alpha & i = 1,2,\cdots,F \\ (1-\beta)^{k-1}\beta & i = F+1,F+2,\cdots,K \end{cases}$$

where $\alpha = (1-Fr_1)/(1-r_1)$ and $\beta = [1-(K-F)r_K]/(1-r_K)$. The fact that T_i is geometrically distributed is intuitive because when the activity levels of those conferees in region L (or R) are the same, the probability that the process leaves region L (or R) does not depend on whom has just finished speaking. The channel holding time Y_{AB} as given by (16) now becomes a geometrically weighted Erlangian distribution which converges to the exponentially distribution as follows:

$$f_{Y_{AB}}(t) = (\alpha\mu)e^{-\alpha\mu t}$$

Similarly, $f_{Y_{BA}}(t)$ is given by:

$$f_{Y_{AB}}(t) = (\beta\mu)e^{-\beta\mu t}$$

Numerical results show that for $a_i \neq a_j$ for a_i's not all equal, $f_{Y_{AB}}(t)$ still looks very much like the exponential distribution with variance differs from that of the exponential distribution by only a few percents.

When the process leaves region L, it will release the channel on link AB or the video source corresponding to the conference concerned goes idle. At the same time one of the conferees in region R will seize a channel on link BA and goes active. The channel idle time on link AB is therefore the same as the channel holding time on link BA. But instead of letting the channel actually goes idle, the network could assign other traffic onto it. As channels are only active half of the time on the average, such sharing of channel resources could lead to a significant increase in the network's traffic carrying capacity. The tradeoff is that a newly activated video source may not get a channel and as a result video-freeze may occur for some conferees. The challenge is then to admit as much traffic as possible while containing the occurrences of video-freeze as well as its duration to a tolerable limit.

IV. Link Level Video-Freeze Statistics

Consider a directed link in the network. Let there be V conferences connected to that link. The link can support at most E simultaneous video transmission (or E channels) without violating video quality constraint. Each conference on the link cycles through active and inactive periods. For convenience, we model the conferences as video sources. Let Y_i and Z_i be i.i.d. random variables representing the length of active and inactive periods, respectively, of video source i. Then, the activity level A_i of source i, i.e. the probability that source i is active is simply

$$A_i = \frac{E[Y_i]}{E[Y_i] + E[Z_i]}$$

Note that A_i is also given by the sum of the activity levels a_j's of the set of conferees on the transmitter side of the directed link. Numerical results show that the two are identical.

When $V \leq E$, channels are always available for active sources and so video-freeze due to lack of channel resources is not possible. When $V > E$, there is a possibility that more than E sources turn active simultaneously on the link and some of them will experience video-freeze.

Let S be a subset of the V sources on the link and S_C be the complement set of S. Let P_S be the steady state probability that only sources in S are active and the others are inactive. Let $|S|$ be the number of elements in S. Since all sources are independent we have

$$P_S = \prod_{i \in S} A_i \prod_{j \in S^C} (1 - A_j)$$

The expected number of frozen video sources $E[W]$ is given as

$$E[W] = \sum_{S:|S|>E} (|S| - E) P_S$$

If a source, say source i, is currently not active, it will turn active with rate $1/E[Z_i]$ where $E[Z_i]$ is the expected length of inactive period. Given that only the sources in S are active, the rate that inactive sources are turning active, denoted as λ_S, is

$$\lambda_S = \sum_{i \in S^C} \frac{1}{E[Z_i]}$$

We define three video-freeze measures as follows:

1. Video-Freeze Probability B: The video-freeze probability B is defined as the ratio of time during which video-freeze occurs to the total time of observation and is given as
$$B = \sum_{S:|S|>E} P_S \tag{17}$$
2. Normalized Average Video-Freeze Duration X: The normalized average video-freeze duration X is defined as the average video-freeze duration of a source normalized by the mean active period of a conferee but under the condition that a video-freeze has occured. It can be obtained by Little's formula as
$$\begin{aligned} X &= \frac{\text{Expected number of frozen video sources}}{\text{rate of sources experiencing video-freeze}} \cdot \frac{1}{\mu^{-1}} \\ &= \frac{\mu E[W]}{\sum_{S:|S| \geq E} \lambda_S P_S} \end{aligned} \tag{18}$$
The average video-freeze duration in seconds is just X multiplied by μ^{-1}.
3. Fractional Video Video-Freeze R: The fractional video-freeze R is defined as the ratio of the total frozen video to the total generated video summing over all video sources on the link. Relating to the state probabilities P_S's, we have
$$\begin{aligned} R &= \frac{\text{Expected number of frozen sources}}{\text{Expected number of active sources}} \\ &= \frac{E[W]}{\sum_S |S| P_S} \end{aligned} \tag{19}$$
This measure is analogous to fractional speech loss in TASI related systems [1].

V. Conference Video-freeze Statistics

Video-freeze statistics derived on a link may not be the video-freeze statistics experienced by the conferees. This is because video may have to pass through several links before reaching a conferee. When a conferee at a particular node, say node C, has terminated the video transmission and another conferee at another node, say node D, starts transmitting, the transmission direction of the path connecting nodes C and D is reversed for that conference (Fig. 2). Let $\Im_{DC}$ be the set of links constituting the path from node D to node C. Assuming that a new video transmission is allowed only when the channel resources on all involved links are available, the new video transmission can be initiated if and only if all the links in $\Im_{DC}$ are free from video-freeze. Let Bi and X_i be the video-freeze probability and normalized average video-freeze duration on link i. Then if we assume the video-freeze statistics are independent for

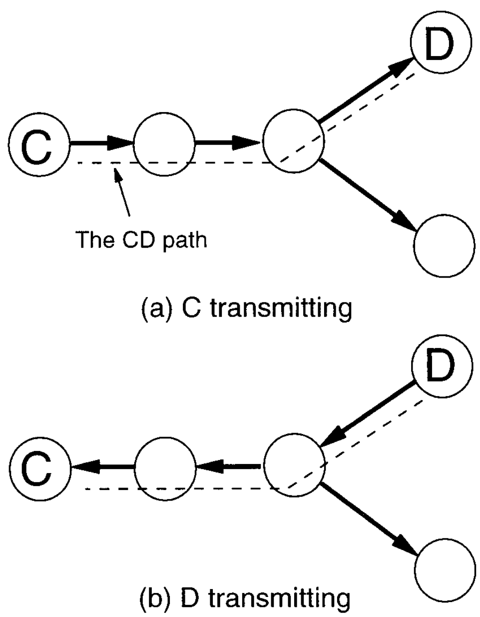

Fig. 2. Change of transmission direction and the set of links involved.

the set of links in $\Im_{DC}$, the video-freeze probability B_{CD} when video transmission is switched from node C to node D is

$$B_{CD} = 1 - \prod_{i \in \Im_{DC}} (1 - B_i) \tag{20}$$

Let Ω be the set of nodes with conferees attached and $\zeta_\omega, \omega \in \Omega$, be the set of conferees homed at node ω. The probability γ_{CD} that video transmission is switched from node C to node D is:

$$\gamma_{CD} = \sum_{i \in \zeta_C} a_i \sum_{i \in \zeta_D} p_{ij}$$

The video-freeze probability B_o experienced by the conferees is obtained by removing the conditioning on the node pair C and D:

$$B_O = \sum_{C,D \in \Omega, C \neq D} \gamma_{CD} B_{CD}$$

If in a switch of video transmission from node C to node D, i links on path CD, say links $L_1, L_2, \cdots, L_i$, are experiencing simultaneous video-freeze, the video-freeze duration $\tilde{V}_{L_1, L_2, \cdots, L_i}$ is given by

$$\tilde{V}_{L_1, L_2, \cdots, L_i} = \max[\tilde{X}_{L_1}, \tilde{X}_{L_2}, \cdots, \tilde{X}_{L_i}] \tag{21}$$

where $\tilde{X}_{L_j}$ is the random variable denoting the video-freeze duration on link L_j. When a video transmission is switched from node C to node D, the normalized average video-freeze duration X_{CD} given a video-freeze is obtained by enumerating all possible simultaneous video-freeze occurrences on the links in $\Im_{DC}$ and weighted them by their respective probabilities of occurrence. These probabilities, denoted as C_i's, are the probability of video-freeze upon a switch of video source and is therefore a call congestion measure [6]. For the derivation of X_{CD} we shall approximate the C_i's by the B_i's or by the time congestion measure given by (17). As noted in [6], the difference between these two measures is usually small. Having that, we can express X_{CD} by

$$X_{CD} = Y_1 / Y_2$$

where

$$\begin{aligned} Y_1 = & \sum_{i \in \Im_{DC}} [B_i \prod_{j \in \Im_{DC}, j \neq i} (1 - B_j)] X_i + \\ & \sum_{i,j \in \Im_{DC}, i \neq j} [B_i B_j \prod_{k \in \Im_{DC}, k \neq i \neq j} (1 - B_k)] \cdot \\ & E[\tilde{V}_{i,j}] + \cdots \end{aligned}$$

$$\begin{aligned} Y_2 = & \sum_{i \in \Im_{DC}} [B_i \prod_{j \in \Im_{DC}, j \neq i} (1 - B_j)] + \\ & \sum_{i,j \in \Im_{DC}, i \neq j} [B_i B_j \prod_{k \in \Im_{DC}, k \neq i \neq j} (1 - B_k)] + \cdots \end{aligned}$$

The first $[\cdot]$ on the numerator is the probability that link i on path DC has a video-freeze, the second $[\cdot]$ is the probability that video-freeze has occurred on both link i and link j on path DC. For B_i small we can assume simultaneous video-freeze on two or more links in $\Im_{DC}$ is very rare, or $B_i B_j \approx 0$. Therefore X_{CD} can be further approximated by

$$X_{CD} \approx \frac{\sum_{i \in \Im_{DC}} B_i X_i}{\sum_{i \in \Im_{DC}} B_i}$$

The normalized average video-freeze duration X_O is given by

$$X_O = \sum_{C,D \in \Omega, C \neq D} \gamma_{CD} X_{CD}$$

Note that fractional video-freeze is only a link level measure. There is no corresponding measure on the conference level.

VI. Illustrative Examples

A. Example 1

In the last section, we have made three assumptions to simplify the derivation of conference level video-freeze measures, namely that the video-freeze statistics are independent from link to link, the B_i's are good substitutes of the C_i's and for B_i small simultaneously video-freeze in two or more links on a path is very rare. In this example, we wish to check these assumptions.

Consider a five node network shown in Fig. 3. Let the number of conferees per conference in the network be 3, 4 or 5 with equal probabilities. Further, let the conferees be randomly located in the network and be connected by a minimum multicast tree (minimum number of edges). Table 1 shows the analytical and simulation results of 10 cases of random conferee distribution. The first column shows

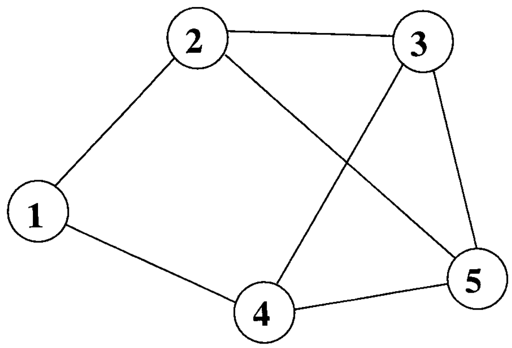

Fig. 3. Network of example 1.

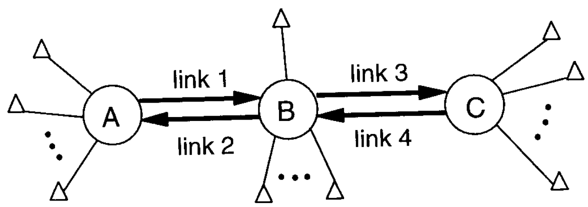

Fig. 4. Network of example 2.

the case numbers and the second column shows the number of admitted conferences. Thus the first two cases both have 26 conferences each but have different randomly generated conferee distributions. The third column shows the probability of video-freeze from equation (20) for a specific two-link path while the fourth column shows the corresponding simulation result. It is seen that the difference is only a few percents, indicating that the approximation used in (20) is a good one for two-link paths.

To verify the second and the third assumptions, we select the three link path from node 1 to node 3 passing nodes 2 and 5. The average video-freeze duration on this path is measured in the simulation runs and compared with that obtained from (22). Table 2 shows the results of 10 cases of random conferee distribution. As there is only a few percent difference between the analytical and the simulation results, the second and the third assumptions appear to be on solid ground.

B. Example 2

In this example, we find the number of conferences that can be accommodated on the 3 nodes network shown in Fig. 4. Let there be E video channels each on link 1 to link 4. For all conferences, let there be one conferee on each of the nodes A, B and C and let the activity level of all conferees be 1/3. $E[Y_i], E[Z_i]$ and A_i are found to be 1.0, 2.0 and 1/3 respectively for all links. Let V be the number of conferences accepted on the network. The video-freeze probability B_i and the normalized average video-freeze duration X_i can be found as:

$$B_1 = B_4 = \sum_{i=E+1}^{V} \binom{V}{i} (\frac{1}{3})^i (\frac{2}{3})^{V-i}$$

$$B_2 = B_3 = \sum_{i=E+1}^{V} \binom{V}{i} (\frac{2}{3})^i (\frac{1}{3})^{V-i}$$

$$X_1 = X_4 = \mu \frac{\sum_{i=E+1}^{V} (i-E) \binom{V}{i} (\frac{1}{3})^i (\frac{2}{3})^{V-i}}{\sum_{i=E}^{V} (\frac{V-i}{E[Z]}) \binom{V}{i} (\frac{1}{3})^i (\frac{2}{3})^{V-i}}$$

$$X_2 = X_3 = \mu \frac{\sum_{i=E+1}^{V} (i-E) \binom{V}{i} (\frac{2}{3})^i (\frac{1}{3})^{V-i}}{\sum_{i=E}^{V} (\frac{V-i}{E[Y]}) \binom{V}{i} (\frac{2}{3})^i (\frac{1}{3})^{V-i}}$$

Fig. 5(a) and (b) show B_i as a function of the number of conferences V on the network. As seen, since the second tandem links (link 2 and link 3) have twice as much traffic as the first tandem links (link 1 and link 4), they have much higher video-freeze probabilities. Also, when $V \rightarrow E, B_i \rightarrow 0$ as it should. Fig. 5(c) and (d) show X as a function of V. It is interesting to note that there is a threshold on V above which X has a very sharp increase.

Given the performance requirement $B_i \leq B^*$, there is a maximum number of video sources V_i^* that can be accommodated on link i. Fig. 5(e) plots the capacity gain factor V_i^*/E of link i for $B^* = 0.01$. A minimum gain of 230% for the first tandem link and 126% for the second tandem link are observed. It is seen that V_1^* is always larger than $V_3 * 3$ because link 1 carries the traffic from node A only whereas link 3 must carry the traffic from both node A and node B. Similar argument can be made when $X_i \leq X^*$ is chosen as the performance requirement (Fig. 5(f)).

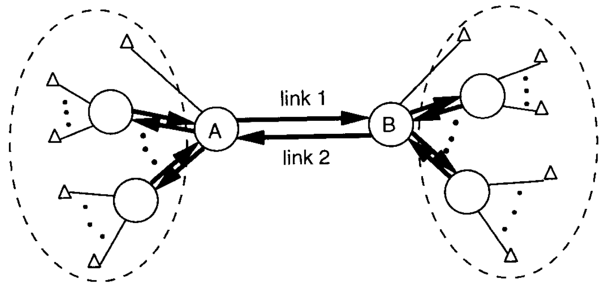

Fig. 6 The multicast tree of example 3.

C. Example 3

In this example, we illustrate the execution of the admission policy. We assume that there are five conferees in all conferences and all conferees have equal activity levels. The

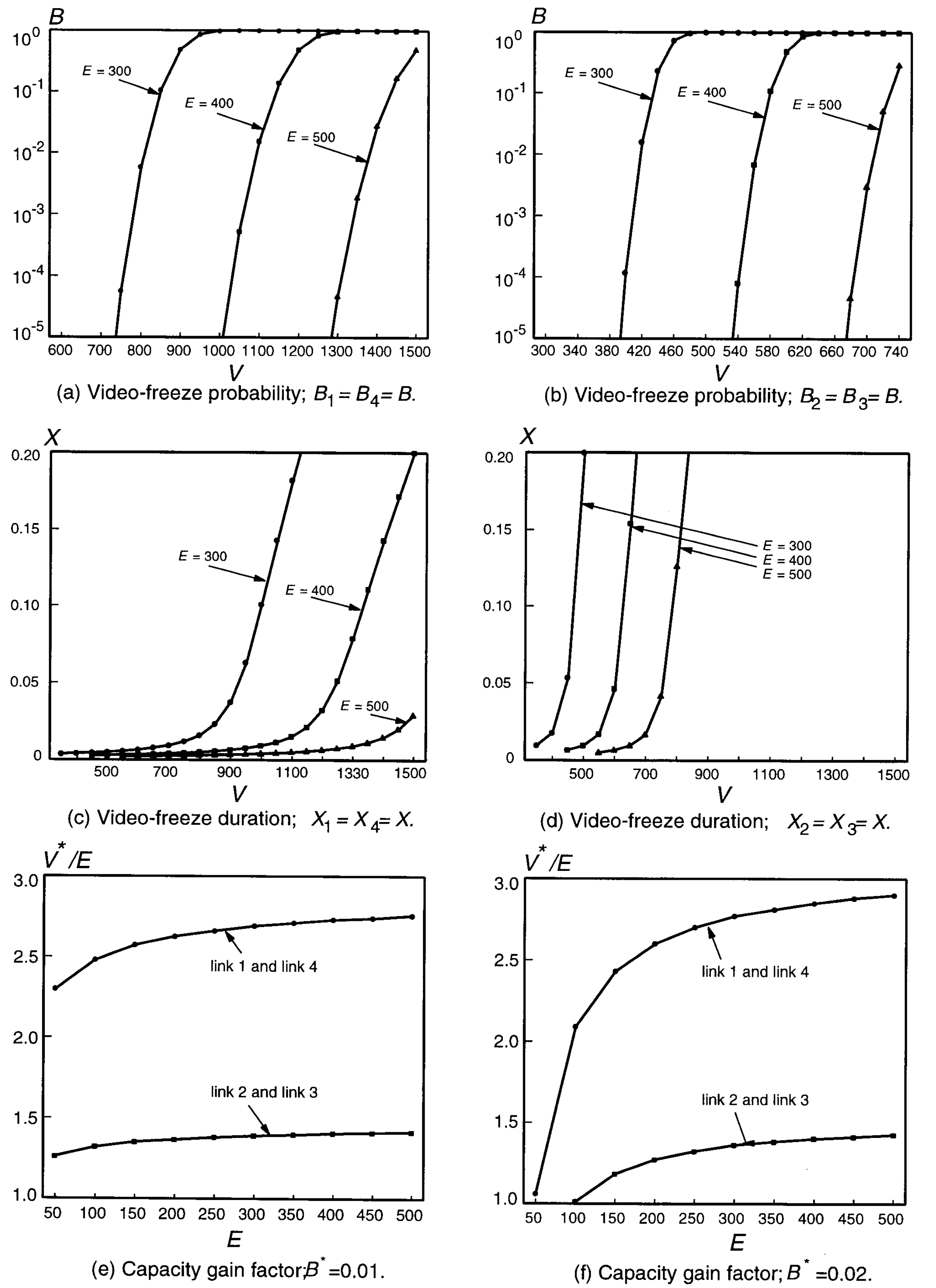

Fig. 5 Results of example 2.

conferees can be randomly located in the network. Consider a particular link pair connecting two adjacent nodes, say node A on the left and node B on the right as shown in Fig. 6. We can identify four types of traffic on the link pair with type i ($i = 1, 2, 3, 4$) having i conferees at the transmitter side of the link and $5 - i$ conferees at the receiver side. Let Q_i be the number of type i conferences on the link pair. The traffic on the link pair can be specified by the quadraple $Q = (Q_1, Q_2, Q_3, Q_4)$. Let the link video-freeze requirement be $B^* = 10-2$. Table 3 shows the video-freeze probability if a new type i conference is admitted. When the link pair is at a load of $Q = (40, 40, 45, 44)$, only types 1, 2 and 3 can be admitted without violating the performance requirement. As a separate case, if $X^* = 0.02$ is chosen as the admission requirement, and the current link load is $Q = (30, 30, 30, 31)$. Table 4 shows that both type 1 and type 2 conferences can be admitted without violating the constraint. Note that if we assume mean active period of a conferee to be 1 minute, the above constraint stipulates that mean video-freeze duration should be smaller than 1.2 seconds.

VII. Conclusion

Video is such a bandwidth hungry traffic that extreme care must be exercised to regulate their use of network resources. We have proposed in this paper a video bandwidth allocation architecture for multipoint multimedia teleconferences on a communication network. By exercising the admission policy, a new conference will be admitted only when all links intended to be used by this new conference satisfy some specified performance requirements. The transmission policy stipulates that only when all links concerned have the required bandwidth will a newly active video source start its transmission. This policy guarantees that ongoing video transmissions will not be degraded due to temporary congestion. A conference traffic model is presented and solved and both link level and conference level video-freeze statistics are derived. Examples illustrating video video-freeze statistics and conference admission decision processes are provided.

References

[1] C.J.Weinstein, "Fractional speech loss and talker activity model for TASI and for packet-switched speech", *IEEE Trans.Commun.*, vol.26, no.8, August 1978.

[2] B.Maglaris, D.Anastassiou, P.Sen, G.Karlsson, J.D.Robbins, "Performance Models of statistical multiplexing in packet video communications", *IEEE Trans.Commun.*, vol.36, no.7, July 1988.

[3] P.Sen, B.Maglaris, N.E.Rikli, D.Anastassiou, "Models for packet switching of variable-bit-rate video sources", *IEEE J. Select. Areas Commun.*, vol.7, no.5, June 1989.

[4] R.Howard, Dynamic Probabilistic Systems, vol.1: Markov Models, p.286, 1971.

[5] H.Kobayashi, Modelling and Analysis: An Introduction to System Performance Evaluation Methodology, pp.108-110, Addison-Wesley, 1978.

[6] M.Schwartz, Telecommunication Networks: Protocols, Modeling and Analysis, p.519, Addison Wesley, 1987.

[7] T.S.Yum, M.S.Chen, Y.W.Leung, "Bandwidth Allocation for Multimedia Teleconferences", *IBM Research Report RC 16271 (#72186)*, November 1990.

Tak-Shing Yum (S'76-M'78-SM'86) received his BSc, MSc, MPh and PhD from Columbia University in 1974, 75, 77 and 78 respectively. He worked in Bell Telephone Laboratories, USA for two and a half years and taught in National Chiao Tung University, Taiwan, for 2 years before joining The Chinese University of Hong Kong in 1982. He has published original research on packet switched networks with contributions in routing algorithms, buffer management, deadlock detection algorithms, message resequencing analysis and multiaccess protocols. In recent years, he branched out to work on the design and analysis of cellular network, lightwave networks and video distribution networks. He believes that the next challenge is designing an intelligent network that can accommodate the needs of individual customers. Peter claims himself to be an existentialist. He enjoys playing bridge and badminton.

Mon-Song Chen received the BSEE degree from National Taiwan University in 1978, the MSEE degree from the University of Washington, Seattle, in 1982, and the PhD degree from Polytechnic Institute of New York, Brooklyn, in 1985. From 1985 to 1990, he was a researcher of high-speed networking architectures, protocol engineering, and WDMA optical networking at IBM's T. J. Watson Research Center. He received several patents in flow and error control designs. From 1990 to 1993, he was the manager of the High Bandwidth Application Group, which develops network-based multimedia technologies and applications. The activities included bandwidth reservation for multiparty video conferences, disk-arm scheduling for multimedia storage management, and the design of a desktop collaboration system called Multimedia Multiparty Teleconference (MMT) system. He is currently the manager of the Global Client/Server Services department in IBM's Application Solution Line of Business.

Yiu-Wing Leung received his BSc and PhD degrees from the Chinese University of Hong Kong in 1989 and 1992 respectively. He is currently a Lecturer in the Department of Computing, Hong Kong Polytechnic, Hong Kong. His current research interests are information networks and large-scale software systems.