Hierarchical Distribution of Video with Dynamic Port Allocation

Tak-Shing P. Yum, Senior Member, IEEE

Abstract—A two-level distribution network for broadcast and interactive video is studied as an example of the hierarchical distribution method. This two-level design has the following advantages.

1) It can replace a large switch by a network of smaller switches, facilitating switch growth and enhancing switch reliability.
2) The second level switches (or the local switches) can be located at convenient places in their respective service districts, reducing the overall circuit mileage of the video distribution system.
3) Video requests can be processed independently by local switches, rendering a large call processor at the central switch unnecessary.

A traffic model for this network is formulated and the optimum capacities of the central and local switches are determined for a given blocking requirement. By adding a small crosspoint switch between the two levels, the output ports of the central switch can be dynamically allocated to the local switches. We show that this sharing of output ports can significantly reduce the size of the central switch.

I. INTRODUCTION

PROPOSALS for residential video networks frequently distinguish two types of video services, broadcast, and interactive. In broadcast mode video signals are transmitted by the network and selected by customers as desired. Individual customers have no control over the content of broadcasts and no ability to dynamically change the network. With interactive video—video conferencing and video on demand [1], for example—network transmission is initiated by customer requests. Interactive video is obviously the preferred service since customers could have some freedom to choose when and what to see. Full-scale interactive video service will be possible with the future introduction of broad-band ISDN. In fact, the main goal of developing broad-band ISDN is to provide switched video service to the subscribers.

Tremendous amount of research has been conducted recently on the distribution of video, including how to accommodate both broadcast and interactive services onto a single system, what kinds of multiplexing and switching technologies could be used and their tradeoffs. Wagner and Menendez [2] reviewed the evolutionary architectures and techniques for video distribution on fibers, from the currently matured analog transmission techniques to fully switched digital systems for broad-band ISDN in the late 1990’s and beyond. Cooperman, Paige, and Sieber [3] described the architecture, the operation and control of a state-of-the-art 64 × 16 broad-band video switch on a chip.

As video distribution technology becomes more and more mature, characterizing service demands and building traffic engineering tools for various design architectures are imminent. In this paper, we choose to do this on a two-level hierarchical distribution network that can accommodate both broadcast and interactive video. The hierarchical architecture has three advantages. First, the total number of switching cells can be reduced. Second, the second-level switch (or local switch) can be located at convenient places in their respective service districts. As a result, the overall circuit-mileage of the video distribution system can be significantly reduced. Third, most of the switching demands can be satisfied independently by the local switches. With this distributed processing feature, the speed and size of the system will no longer be limited by the speed of the central call processor.

Specifically, for this two-level distribution network we shall first formulate a traffic model and derive the blocking probabilities of both broadcast and interactive video services. We then formulate the optimum dimensioning of the central and local switches as a nonlinear integer program. By adding a small subsidiary switch the output ports of the central switch can be dynamically shared by the array of local switches. We shall show by numerical examples that such sharing can reduce the size of the central switch without increasing the blocking probability of video demands.

II. SYSTEM ARCHITECTURE

Fig. 1 shows the details of a two-level distribution system. It consists of a central switch and K local switches. The central switch is connected to a broad-band backbone network where the broadcast video programs are supplied and the circuits for interactive video are allocated. It has E input ports and M (M ≥ E) output ports and has embedded in it a copy network [4]. It is able to transmit a maximum of E video programs simultaneously to the M output ports, with some of them receiving the same programs. The M output ports supply the video programs to the K local switches, which in turn duplicate and distribute them to the subscribers in their respective regions. Let all local switches have J input ports and N output ports. Then the entire switching system can serve a maximum of \( N_c = NK \) subscribers.
Let us first consider the distribution of broadcast video. After a customer turns on his TV set, he will select a particular TV program from his remote control unit. His selection will be transmitted to the local switch through an embedded signaling channel. If the local switch is currently transmitting that program, it just sends a copy to that customer. On the other hand, if that customer is the first one in that region to request that program (after a period of idling), the local switch sends a signal to the central switch to ask for a copy and passes it on to the customer. Subsequent requests for the same program can be satisfied by the local switch through electronic copying. If on the other hand, a customer wants to start an interactive video session, his/her request will be relayed through the local switch to the backbone network and a dedicated circuit is allocated to the two communication parties.

III. TRAFFIC MODEL AND BLOCKING PROBABILITIES

Let there be a total of $C$ broadcast programs available in the entire region of service and let $c_i$ ($i = 1, 2, \ldots, C$), be the probability that program $i$ is being watched. For simplicity of computation, let us reorder the program numbers such that $c_1 \geq c_2 \geq c_3 \cdots \geq c_C$. A typical distribution is shown in Fig. 2.

At the busiest hour of a day let $p_1$ be the probability that a subscriber chooses to watch a broadcast program and $p_2$ be the probability that a subscriber requests an interactive session. Then in a region with $N$ subscribers, the probability that there are $n_1$ broadcast requests and $n_2$ interactive requests is

$$A(n_1, n_2) = \binom{N}{n_1, n_2} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{N-n_1-n_2}$$

$$= \frac{N!}{n_1! n_2! (N-n_1-n_2)!} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{N-n_1-n_2}$$

$$0 \leq n_1 + n_2 \leq N. \quad (1)$$

The respective marginal probabilities are denoted as $A(n_1, \cdot)$ and $A(\cdot, n_2)$.

The $n_2$ interactive video sessions require precisely $n_2$ channels. But many of the $n_1$ subscribers on broadcast programs will be sharing channel if they are watching the same program. Thus, given $n_1$, the probability that $r$ broadcast channels are required $b_r(n_1)$ is

$$b_r(n_1) = \sum_{\Omega_1} \sum_{\Omega_2} \binom{n_1}{m_1, m_2, \ldots, m_r} c_1^{m_1} c_2^{m_2} \cdots c_r^{m_r}$$

$$r = 1, 2, \ldots, \min(n_1, C)$$

where $c_i^{m_i}$ is the probability that $m_i$ customers are watching program $k_i$ ($i = 1, 2, \ldots, r$),

$$\Omega_1 = \{(k_1, k_2, \ldots, k_r) | 1 \leq k_1 < k_2 < \cdots < k_r \leq C\}$$

is the set of all the combinations of $r$ programs selected from the set of $C$ programs, and

$$\Omega_2 = \{(m_1, m_2, \ldots, m_r) | m_1 + m_2 + \cdots + m_r = n_1\}$$

is the set of all the combinations of the $n_1$ customers watching the $r$ programs specified by the elements of $\Omega_1$.

Thus, given $n_1$ and $n_2$, the probability that a total of $j$ channels are required, denoted as $d_j(n_1, n_2)$, is just the probability that $j - n_2$ channels are required by the broadcast video. Therefore,

$$d_j(n_1, n_2) = \begin{cases} 
0 & 0 \leq j \leq n_2 - 1 \\
\frac{b_{j-n_2}(n_1)}{n_2 \leq j \leq n_2 + \min(n_1, C)} & n_1 + n_2 \geq j > 1 
\end{cases} \quad (2)$$

with $d_0(0, 0) = d_1(1, 0) = d_1(0, 1) = 1$. Removing conditioning on $n_1$ and $n_2$, we have

$$d_j = \operatorname{Prob}[j \text{ channels are required at the local switch}]$$

$$= \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} d_j(n_1, n_2) A(n_1, n_2)$$

$$j = 0, 1, 2, \cdots, N. \quad (3)$$

A. Blocking of the Interactive Video

Let there be $J$ input ports at all local switches. Then given $n_1$ and $n_2$ the probability that a requested channel is blocked...
at a local switch, denoted as $B_L(J|n_1, n_2)$, is

$$B_L(J|n_1, n_2) = \sum_{j=1}^{n_2 + \min(n_1, C)} \frac{j - J}{j} d_j(n_1, n_2).$$

Removing conditioning on $n_1$ and $n_2$, we have

$$B_L(J) = \sum_{n_2=0}^{N} \sum_{n_1=0}^{N-n_2} B_L(J|n_1, n_2) A(n_1, n_2).$$

Next, let $W_i$ be the number of requests for interactive video channels from local switch $i$ to the central switch. Then, for $i = 1, 2, \ldots, K$ and $w = 0, 1, 2, \ldots, J$, we have

$$P[W_i = w] = P[W_i = w, j \leq J] + P[W_i = w, j > J]$$

$$= \sum_{j=0}^{J} \sum_{n_1=0}^{N} d_j(n_1, w) A(n_1, w)$$

$$+ \sum_{j=J+1}^{N} \sum_{n_2=0}^{N-n_2} \sum_{n_1=0}^{N-n_2} \binom{n_2}{j} \binom{n_1}{J} \cdot d_j(n_1, n_2) A(n_1, n_2).$$

Let $W_S = W_1 + W_2 + \cdots + W_K$ be the total number of interactive video requests from all $K$ regions. Since all $W_i$'s are i.i.d. random variables, the distribution of $W_S$ is just the $K$-fold convolution of the distribution of $W_1$ with itself. Assuming all $C$ broadcast programs are available at the central switch, the blocking probability $B_C(E, J, K)$ for the interactive video at the central switch is

$$B_C(E, J, K) = \sum_{k=E-C+1}^{KJ} \frac{k - (E - C)}{k} P[W_S = k].$$

The overall blocking probability of interactive video $B_O(E, J, K)$ given that there are $E$ input ports at the central switch and $J$ input ports at each of the $K$ local switches is

$$B_O(E, J, K) = B_L(J) + B_C(E, J, K).$$

B. Blocking of the Broadcast Video

Blocking of broadcast video occurs when the selected program is not currently being broadcast to the local switch and all $J$ input ports of the local switch are occupied. The blocking probability therefore depends on the particular program being selected. Consider the selection of program $z$. Given $n_1$ (as defined before), the probability that program $z$ is not in a set of $j$ broadcast video programs, denoted as $f_j(z, n_1)$, is

$$f_j(z, n_1) = \sum_{S_j} \left[ \frac{1}{\binom{n_1}{m_1, m_2, \ldots, m_j} \cdot e_{k_1} \cdot e_{k_2} \cdots e_{k_j}} \right]$$

$$j = 1, 2, \cdots, C - 1.$$

where

$$S_j = \{(k_1, k_2, \ldots, k_j) | 1 \leq k_1 < k_2 < \cdots < k_j \leq C; \quad \text{all } k_i \neq z \}$$

is the set of all combinations of $j$ programs selected from the set of $C - 1$ programs (excluding program $z$) and

$$Y_j = \{(m_1, m_2, \ldots, m_j) | m_1 + m_2 + \cdots + m_j = n_1 \}$$

is the set of all combinations of the $n_1$ customers watching the specified $j$ programs. The probability that program $z$ is blocked, denoted as $B_B(z)$, is obtained by removing the conditionings of $j$ and $n_1$ on $f_j(z, n_1)$:

$$B_B(z) = \sum_{n_1=1}^{N} \sum_{j=1}^{\min(n_1, C)} f_j(z, n_1) A(n_1, J - j). \quad (4)$$

It is easy to imagine that the blocking probability is practically zero for popular programs and is always smaller than the blocking probability of the interactive video for infrequently accessed programs. Blocking requirement therefore needs only be set on the interactive video for switch capacity optimization.

IV. CAPACITY OPTIMIZATION

Aside from the advantages of distributed processing, the main reason for adopting the two-level architecture is that the switch complexity can be reduced. In this section, we derive the minimum complexity configurations for the crosspoint and the broadcast banyan switches.

A. Crosspoint Switch

The crosspoint switch can perform the multicast function by making multiple output connections. Fig. 3 shows the switch connections whereby program 1 is being received by output ports 3 and 6, program 2 is being received by output ports 1, 2, and 8, etc. The complexity of a crosspoint switch is usually measured by the total number of crosspoints $L$ in the switch. For the switching system shown in Fig. 1, $L$ is given as

$$L = E K J + K J N. \quad (5)$$

The minimum complexity configuration of the switching system can be determined by minimizing $L$ with respect to $E$, $J$, and $K$ simultaneously subject to $B_S(E, J, K) \leq B^*$ where $B^*$ is given blocking requirement. A popular method of solving this type of nonlinear integer program is by Branch and Bound [10].
In comparison, if a single crosspoint switch is used and the same blocking requirement $B^*$ is to be satisfied, a minimum of $E_o$ input ports are needed where $E_o$ is given by

$$E_o = \min \left[ E - \sum_{n_2 = E-C}^{N_o} \frac{n_2 - (E-C)}{n_2} A(n_2, n_2) \leq B^* \right]. \quad (6)$$

The total number of crosspoints $L_o$ in this case is just $L_o = E_o N_o$.

B. Broadcast Banyan Switch [4], [6]–[9]

The same optimization problem can be formulated with the use of a broadcast banyan switch. Here, the complexity is measured in terms of the total number of switching units. The central switch can be partitioned into four parts, namely, the running adder, the copy network, the Batcher network, and the banyan network. Their complexities are shown in Fig. 4. For such a switch with $E$ inputs and $M$ outputs, we define the switch complexity function $\phi(E, M)$ as

$$\phi(E, M) = \frac{M}{4} (\log M)^2 + \frac{M}{2} \log M + \frac{M}{2} \log E. \quad (7)$$

Since the $K$ local switches are similar to the central switch except that the number of input and output ports are $J$ and $N_o/K$, the complexity $L$ of the switching system is

$$L = \phi(E, J K) + K \phi(J, N_o/K). \quad (8)$$

As before, the optimum $E$, $J$, and $K$ values can be determined by minimizing $L$ subject to $B_o(E, J, K) \leq B^*$. The single stage switch has a complexity $L_o$ given by $L_o = \phi(E_o, N_o)$ where $E_o$ is given by (6).

V. DYNAMIC PORT ALLOCATION

The previous design requires $KJ$ output ports in the central switch. But since it is highly unlikely that all $K$ local switches will have all their input ports loaded simultaneously, using $KJ$ output ports at the central switch is overengineering.

But if a smaller number of output ports is to be used, how are they going to be shared? In the following, we propose the use of a simple intermediate crosspoint switch, called the port allocation (PA) switch, to perform the dynamic sharing function.

Fig. 5 shows a PA switch with its connections to the central and local switches. Among the $J$ input ports of the local switch, let us assign $x$ of them to be the contention ports. These $x$ ports are connected to the PA switch to share, together with the contention ports of the other local switches, the $y$ common-shared output ports of the central switch. The other $J-x$ ports of the local switch are called the dedicated ports and are connected directly to the central switch. Thus, the central switch has a total of $(J-x)K + y$ output ports and the PA switch has $y$ input ports and $Kx$ output ports.

Video calls from a local switch are normally assigned to the dedicated ports. Only when all dedicated ports are occupied will the traffic be overflowed to the contention ports. Under a fairly rare occasion that

1. the $y$ common-shared output ports are all occupied,
2. there are one or more free dedicated ports in a local switch, and
3. there are one or more occupied contention ports in the same switch,

a video session on a contention port will be reassigned to one of the dedicated port. In doing so, one of the common-shared output ports of the central switch can be freed to serve new calls from the contention ports of the local switches.

For the calculation of the blocking probability, we shall assume in the following that $N$ is sufficiently large such that all $C$ broadcast programs are being distributed in all local switches. Thus, the number of input ports that can be used for interactive video at each local switch is $J-C$. The number of interactive video connections in local switch $i$, denoted as $W_i$,
has distribution
\[ P[W_i = k] = \binom{N}{k} p_2^k (1 - p_2)^{N-k}. \]

Given \( W_1 = k_1, W_2 = k_2, \ldots, W_K = k_K, \) the blocking probability \( B(x, y, J, K, E) \) of the interactive video programs consists of three terms, \( B_1, B_2, \) and \( B_3 \) denoting the blocking contributions from the local switches, the PA switch and the central switch, respectively. Let \( a_i \) be the number of interactive video requests that gets blocked at local switch \( i \) or
\[ a_i = \max[k_i - (J - C), 0] \]
and \( b \) be the number of requests that overflow onto the PA switch, or
\[ b = \sum_{i=1}^{K} \min[\max[k_i - (J - C - x), 0], x]. \]

Furthermore, let
\[ a_s = \sum_{i=1}^{K} a_i, \]
\[ k_s = \sum_{i=1}^{K} k_i. \]

Then
\[ B_1 = \frac{a_s}{k_s}, \]
\[ B_2 = \frac{\max[b - y, 0]}{b}, \]
\[ B_3 = \frac{\max[k_s - a_s - \max[b - y, 0] - (E - C), 0]}{k_s - a_s - \max[b - y, 0]}. \]

Removing conditioning on the set of \( k \)'s, we have
\[ B(x, y, J, K, E) = \sum_{k_1=0}^{N} \cdots \sum_{k_K=0}^{N} \left( (B_1 + B_2 + B_3) \prod_{i=1}^{K} P[W_i = k_i] \right). \]  

Let the cost of a switching unit in the broadcast banyan switch be one and that of a crosspoint in the PA switch be \( \alpha \) (\( \alpha < 1 \)). The search of the optimal switch configuration can be formulated as a nonlinear integer program as follows. (The formulation for the crosspoint switch is similar.)
\[
\min_{x, y, J, K, E} L = \phi(E, (J - x)K + y) + K\phi(J, N_0/K) + \alpha Kxy
\]
subject to
1) \( B(x, y, J, K, E) \leq B^* \)
2) \( x, y, J, K, E \) positive integers.

The central switch complexity reduction ratio \( \gamma \) is
\[ \gamma = \frac{\phi(E_t, J_1 K_1) - \phi(E_t, (J_1 - x)K_1 + y) - \alpha K_1 xy}{\phi(E_t, J_1 K_1)}. \]

VI. EXAMPLES AND DISCUSSIONS

As an example of optimal design, consider a network serving a population of \( N_o = 1000 \) users. Let \( C = 10, B^* = 10^{-3}, p_1 = 0.5, \) and \( p_2 = 0.1. \) With the use of crosspoint switches in hierarchical distribution the optimal values of \( K, J, E, \) and \( L \) are \( K^* = 9, J^* = 29, E^* = 130, \) and \( L^* = 62930. \) If a single switch is used, \( L_o = 128 \) 000 under the same blocking requirement with \( E = 128. \) Thus, using hierarchical distribution, the switch complexity is reduced to 49% of the original. This percentage gets smaller and smaller as \( N_o \) increases. Thus, for \( N_o = 2000, \) the complexity is reduced further to 38% with \( K^* = 15, J^* = 32, E^* = 237, \) and \( L^* = 177760. \) To investigate the sensitivity of the optimal design to \( K, \) we plot, in Fig. 6, \( L \) versus \( K. \) Here for each \( K \)
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>( N_0 = 1000 )</th>
<th>( N_0 = 2000 )</th>
<th>( N_0 = 1000 )</th>
<th>( N_0 = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( X^* )</td>
<td>( Y^* )</td>
<td>( L^* )</td>
<td>( E^* )</td>
</tr>
<tr>
<td>PA</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>32</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>163 819</td>
<td>13 148</td>
<td>27 764</td>
<td></td>
</tr>
<tr>
<td></td>
<td>237</td>
<td>134</td>
<td>246</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>24</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J^* )</td>
<td>29</td>
<td>32</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>( K^* )</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( L^* )</td>
<td>177 760</td>
<td>13 532</td>
<td>28 407</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>128</td>
<td>234</td>
<td>128</td>
<td>234</td>
</tr>
<tr>
<td>Stage</td>
<td>128 000</td>
<td>468 000</td>
<td>18 120</td>
<td>42 581</td>
</tr>
</tbody>
</table>

Fig. 6. Sensitivity of the optimal design to \( K^* \), crosspoint switch.

Fig. 7. Sensitivity of the optimal design to \( K^* \), broadcast banyan switch.

chosen, \( L \) is optimized with respect to \( J \) and \( E \). The optimal \( J \) and \( E \) values for specific \( K^* \)'s are also shown.

Using the broadcast banyan switch, \( K^* = 13, J^* = 24, E^* = 134, \) and \( L^* = 13 532 \). In comparison, \( L_o = 18 036 \) under the same blocking probability. Thus, the complexity is reduced to 75% for this case. When \( N_0 \) is increased from 1000 to 2000, the complexity reduction percentage is reduced from 75 to 66.6% with \( K^* = 21, J^* = 26, E^* = 246, \) and \( L^* = 28 407 \). The optimal single stage configuration has \( L_o = 42 581 \) and \( E = 234 \). Fig. 7 shows the complexity of the optimal configuration as a function of \( K^* \).

The optimal design with the addition of the PA switch is shown in Table I. To summarize, using the crosspoint switch \( \gamma = 9.8\% \) and \( \gamma = 12.3\% \) for \( N_0 = 1000 \) and \( N_0 = 2000 \), respectively. For the broadcast banyan switch \( \gamma \) depends on \( \alpha \), which is difficult to determine. Using a hypothetical value of \( \alpha = 0.3, \gamma = 8.4, \) and \( \gamma = 7\% \) are obtained for \( N_0 = 1000 \) and \( N_0 = 2000 \), respectively. Note that \( \gamma \) is positive even for \( \alpha = 1 \).

VII. CONCLUSION

Video will constitute the bulk of the traffic in broad-band ISDN. It is therefore essential to understand the switching requirements for this traffic. A video traffic model and an optimal two-level hierarchical video distribution network are studied in this paper. Reduction of switch size via dynamic port allocation is suggested. Numerical results show that the hierarchical configuration can offer substantial reduction of switch complexity and that the addition of a port allocation switch can further reduce the complexity of the central switch by 10%. Extensions of the present work include the optimal design of switch configuration for multirate channels and the dynamic management of channel resources.

ACKNOWLEDGMENT

This research was carried out while the author held a visiting fellowship at Bellcore, U.S.A. during the summer of 1989. It is a pleasure to acknowledge the invitation offered by E. Nussbaum and the helpful conversations the author had with T. T. Lee. Sincere thanks are also due to E. W. M. Wong who pointed out an error in my blocking probability derivation, to M. Chen and Y. -W. Leung who developed the optimization programs, and to S. H. Kan who typed the manuscript. During the revision of this paper, the reviewers had suggested many changes that lead to a smoother presentation.
REFERENCES


Tak-Shing P. Yum (S’76–M’78–SM’86) for a photograph and biography, see p. 1140 of the July 1991 issue of this TRANSACTIONS.