Design issues for mixed media packet switching networks

by D. HUYNH, H. KOBAYASHI** and F. F. KUO†

University of Hawaii
Honolulu, Hawaii

ABSTRACT

In this paper we present some of the important design issues for packet switching networks with both satellite and terrestrial components—which we call mixed media packet switching networks. Satellite packet switching has considerable promise for low cost, high bandwidth data communications. However there is inherent high delay in satellite links which does not appear in ground links. Therefore a mix of the two communications media offers the advantages of low-cost/high bandwidth together with low-delay communications where required. In this paper we examine a number of tradeoffs which offer guidelines for the design and optimum utilization of mixed media networks.

INTRODUCTION

In 1968 the Advanced Research Projects Agency (ARPA) of the U. S. Department of Defense began implementation of a computer-communication network which permits the interconnection of heterogeneous computers at geographically distributed centers throughout the United States. This network has come to be known as the ARPANET, and has grown from the initial four node configuration in 1969 to almost forty nodes (including satellite nodes in Hawaii, Norway, and London) in late 1974. The major goal of ARPANET is to achieve resource sharing among the network users. The resources to be shared include not only programs, but also unique facilities such as the powerful ILLIAC IV computer and large global weather data bases that are economically feasible when widely shared.

The ARPANET employs a distributed store-and-forward packet-switching approach that is much better suited for computer-communication networks than the more conventional circuit-switching approach. Reasons favoring packet switching include lower cost, higher capacity, greater reliability and minimal delay. The CCITT (Comité Consultatif International Télégraphique et Téléphonique), an international standards organization, defines a packet as “a group of binary digits including data and call control signals which is switched as a composite whole. The data, call control signals and possibly error control information are arranged in a specific format.”

Multi-access satellite channels

Communication satellites operating in packet mode are becoming increasingly important for consideration in the design of large computer communication networks. In particular, most possess the following characteristics which are of special importance: the satellite’s antenna coverage allows any of a large number of ground stations to access it at any time (multi-access), and its transmissions can be received by all of these stations at all times (broadcast).

Up to the present, packet switched computer networks have mainly utilized terrestrial communications links. Recently, Telenet Communications Corporation, one of the new value-added carriers announced plans to offer public packet switched data service in which terrestrial and satellite links will be available. ARPA has also planned to augment its terrestrial links with satellite communications and has commissioned several satellite IMPs or SIMPs built by Bolt Beranek and Newman Inc. (BBN). In mid 1975, BBN and the British Post Office jointly conducted a multi-access satellite experiment using the Atlantic INTELSAT IV satellite with SIMPs located at the ETAM ground station in the U.S. and the Goonhilly Downs ground station in England.

The basis of this experiment is the work on the
ALOHANET of THE ALOHA SYSTEM project at
the University of Hawaii. The multiplexing tech­
nique that is used by the ALOHANET is a multi­
access packet switching method that has come to be
known as the ALOHA technique.

Based upon the ALOHA multiplexing method or
variations thereof, a number of techniques have been
proposed for the utilization of a satellite channel in a
packet-switched data network in a way which allows
all stations to dynamically share the channel capac­
ity. All are based on the division of the channel into
time slots approximately equal to a packet
transmission time. The approach described by Roberts
and Kleinrock and Lam in References 12 and 13 makes
use of a technique called “slotted ALOHA random
access” or “slotted ALOHA.” The channel capacity
of a slotted ALOHA channel was estimated to be
1/e = 36%.

Goals of this paper

In this paper we will examine a number of key issues
in the design of a packet-switched communication net­
work composed of a terrestrial store-and-forward
packet switching component combined with a multi­
access/broadcast satellite as depicted in Figure 1.

The factors involved in the design optimization of a
packet switched network are:

- Node location and traffic matrix
- Topology of links
- Capacity of links
- Routing
- Flow control, other network design, etc.

To attempt to optimize any one factor, such as the
capacity of links, it is necessary to assume that the
factors higher in the list must be given. However, to
decide optimum choices for the factors high on the list
may require detailed calculations of all the lower ones.
Fortunately, such factors as node location are usually
not design variables, being fixed by practical considera­
tions. Frank, Frisch and Chou have developed sub­
optimum procedures for the topological layout of a
packet switched network. Thus we will concentrate on
the following problems in our paper:

1. Routing of packets via ground or satellite
2. Capacity assignments for ground and satellite
   channels
3. Retransmission strategies

In this paper we will concentrate on the discussion of
design issues such as throughput, delay, cost, etc.,
rather than dwell on the theoretical development of
the design equations which is given in a companion
paper to be published elsewhere.

THE NETWORK MODEL

The network model under consideration consists of
a terrestrial store-and-forward packet switching net­
work referred to here as the ground subnet, and a
multi-access/broadcast satellite which together with
the SIMP ground stations is the satellite subnet.

For the ground subnet we will use the model given
in the paper by Crowther, et al., and will reproduce
the succinct definitions of network terminology given
in that paper:

Nodes—The nodes of the network are real-time
computers, with limited storage and processing re­
sources, which perform the basic packet-switching
functions.

Hosts—The Hosts of the network are the com­
puters, connected to nodes, which are the providers
and users of the network services.

Lines—The lines of the network are some type of
communications circuit of relatively high bandwidth
and reasonably low error rate.

Connectivity—We assume a general, distributed
topology in which each node can have multiple paths
to other nodes, but not necessarily to all other nodes.
Simple networks such as stars or rings are degen­
erate cases of the general topology we consider.

Message—The unit of data exchanged between
source Host and destination Host.

Packet—The unit of data exchanged between adja­
cent nodes.

Acknowledgment—A piece of control information
Design Issues for Mixed Media Packet Switching Networks

Every region has a SIMP and the IMPs in that particular region can only go through one and only one SIMP. The regionalization is determined by the closeness of an IMP to a SIMP in terms of the number of hops and the distance between them. Such a structure is shown in Figure 1.

ROUTING

With a mixed media network the issue of routing is a major concern. With two possible courses to choose from—one via satellite and one via ground, the issue is to choose the set of routes so as to minimize the overall average network delay. The tradeoffs to consider are these: the satellite channel has an inherent minimum delay of .26 seconds. However, satellite capacity is less costly than ground channel capacity for medium to long distances, and therefore more satellite capacity is available at less cost than comparable facilities on the ground. The ground channels are inherently faster than the satellite channel, but because of capacity limitations, are subject to queuing delays, which combined with the store-and-forward nodal processing delays may, in heavy traffic situations, result in larger overall delays than the satellite delays.

To summarize, satellite channels have greater delays but also more cost effective channel bandwidth than ground channels.

The routing procedure used in the ARPANET is a distributed adaptive algorithm in which each node has a routing table which is periodically updated with minimum distance estimates from its immediate neighbors. Unfortunately the distributed adaptive algorithm is extremely difficult to describe analytically and could not be applied to our analytical model. Instead we have chosen a deterministic split traffic (bifurcated) routing strategy which because it allows traffic to flow on more than one path between a given source-destination node pair gives a better balance than a fixed routing procedure. It should be noted that our analytical model does not require any specific routing algorithm, but can accommodate any routing algorithm that can be modelled mathematically.

Optimal routing of packets

Using the network model of Figure 1 the main problem in packet routing is to route packets from one region to another via either ground or satellite links in such a way as to minimize the overall average delay of the network. Given the topology of the network and the capacities of the ground and satellite links, we assume a demand matrix \([v_{ij}]\) where \(v_i\) is the average Poisson input in packets/sec from node \(i\) to node \(j\). Let us define the routing index \(g_{ij}\) as the fraction of \(v_i\) sent through the ground net and let \(g_{ij}=1-g_{ij}\) be the fraction of \(v_i\) sent through the satellite net by first routing.
to the SIMP in the region where IMP \( i \) resides, then transmitting through the satellite channel to the SIMP in the region where IMP \( j \) resides. As a result, \( g_{ij} = 1 \) if both IMPs \( i \) and \( j \) are in the same region.

Let us define \( T \) as being the overall average delay of the network, averaged over all source-destination node pairs and all routes dictated by the routing algorithm. The optimal routing strategy for mixed media networks can be stated as:

\[
\begin{align*}
\text{min} & \quad T; \text{ subject to } \{0 \leq g_{ij} \leq 1 \text{ for all } i,j\} \\
& \text{g}_{ij} \\
\end{align*}
\]

For our analysis, we define the following notation:

\[
\gamma = \sum_{i,j} g_{ij} \quad \text{total traffic in the net}
\]

\[
\tau_{ij} = \text{average delay for a packet travelling from IMP } i \text{ to IMP } j
\]

\[
\sigma(i) = \text{index of the region (hence that of the corresponding SIMP) to which IMP } i \text{ belongs}
\]

\[
\tau_{s} = \text{average delay for a packet transmitted from one SIMP to another via the satellite channel}
\]

\[
T_{ij} = \text{overall average delay for a packet travelling from IMP } i \text{ to IMP } j \text{ (averaged over ground and satellite links)}
\]

We can readily derive the following relationships:

For the ALOHA scheme

\[
T_{ij} = g_{ij}\tau_{ij} + \overline{g}_{ij}(\tau_{ui}(i) + \tau_{s} + \tau_{f}(j))
\]

and

\[
T = \frac{1}{\gamma} \sum_{i,j} \gamma_{ij} T_{ij}
\]

\[
= \frac{1}{\gamma} \sum_{i,j} \left\{ g_{ij}\tau_{ij} + \overline{g}_{ij}(\tau_{ui}(i) + \tau_{s} + \tau_{f}(j)) \right\} + \frac{1}{\gamma} \sum_{i,j} \overline{g}_{ij}\tau_{s}
\]

(2)

For the MASTER scheme, in order to obtain an expression analogous to Equations (1) and (2), we further introduce the notation:

\[
P_{s} = \text{probability that a packet transmitted from SIMP } \sigma \text{ into the satellite channel is successful}
\]

Then we obtain

\[
T_{ij} = g_{ij}\tau_{ij} + \overline{g}_{ij}[(\tau_{ui}(i) + P_{s(i)}\tau_{ui}(j)) + (1-P_{s(i)})\tau_{ui}(i)]
\]

and

\[
T = \frac{1}{\gamma} \sum_{i,j} \gamma_{ij} T_{ij}
\]

\[
= \frac{1}{\gamma} \sum_{i,j} \left\{ g_{ij}\tau_{ij} + \overline{g}_{ij}[(\tau_{ui}(i) + P_{s(i)}\tau_{ui}(j)) + (1-P_{s(i)})\tau_{ui}(i)] \right\} + \frac{1}{\gamma} \sum_{i,j} \overline{g}_{ij}\tau_{s}
\]

In Equations (2) and (4) the terms with \( g_{ij} \) represent the portion of traffic from IMP \( i \) to IMP \( j \) routed entirely through the ground net and those terms with \( \overline{g}_{ij} \) represent the portion routed from node \( i \) to the regional SIMP \( \sigma(i) \), from \( \sigma(i) \) to \( \sigma(j) \), the regional SIMP in which IMP \( j \) resides and then from \( \sigma(j) \) to IMP \( j \).

Thus even in the satellite routing, there is some associated ground traffic to get to and from the regional SIMPs. Our mathematical model is flexible enough to permit SIMPs at every node, and if this were the case, obviously the \( \tau_{ui}(i) \) and \( \tau_{ui}(j) \) terms would be zero. The term \( \tau_{s} \) is derived from the finite population model of a slotted ALOHA channel as originally given in Lam \(^{12}\) and in a slightly different form by the authors.\(^{17}\)

Let \( \lambda_{i} \) be the traffic on link \( i \). We can derive \( \lambda_{i} \) as a function of \( g_{ij} \), the traffic demand \( \gamma_{ij} \) and the routing algorithm. Since the equation is fairly complex we will not give it here but refer the reader to a companion paper for its derivation.\(^{17}\) Given \( \lambda_{i} \) and neglecting the delay due to transmission errors and nodal processing and ground propagation time, the average delay along any ground link can be described by the \( M/M/1 \) queuing model by virtue of the well-known independence assumption and Jackson's decomposition theorem.\(^{20}\)

\[
T_{l} := \frac{1}{\mu C_{l} - \lambda_{l}}
\]

(5)

where \( C_{l} \) is the capacity of channel \( l \) and \( \frac{1}{\mu} \) the average packet length on that channel. The overall average delay can be expressed in terms of the link traffic \( \lambda_{l} \) and link delays \( \tau_{s} \) (\( T_{l} \)) as

\[
T = \frac{L}{\gamma} \left[ \sum_{l=1}^{L} \lambda_{l}T_{l} + \lambda_{s}\tau_{s} \right]
\]

(6)

where \( L = \text{total number of links in the net} \), and \( \lambda_{s} \) the total traffic rate on the satellite channel, i.e.,

\[
\lambda_{s} = \sum_{l=1}^{L} \sum_{ij} g_{ij}\gamma_{ij}
\]

We can show that \( T \) in Equation (6) is convex with respect to \( g_{ij} \) so that we can obtain a minimum overall delay by varying \( g_{ij} \) according to a selected optimization algorithm. The optimum choice of \( g_{ij} \) depends also upon the topology of the network, the location of the SIMPs, the capacities of the ground and satellite links, the specific terrestrial routing algorithm, as well as the traffic demand matrix \( \{\gamma_{ij}\} \), all of which must be specified in advance of the optimization computation.

Since the optimum selection of \( g_{ij} \) is not critically dependent upon a particular optimization procedure, we used the algorithm known as the BOX COMPLEX procedure\(^{28}\) since this is one of the optimization algorithms which does not require derivatives of the objective function.

As an example, consider the network model in Figure 2 in which the two regions consist of nodes \((1,2,3,4)\) and nodes \((5,6,7,8)\), and the regional SIMPs are located at nodes 1 and 7. The traffic demand matrix assumed to be uniform with \( \gamma_{ij} = 20 \text{ packets/sec.} \), \( \forall \end{enumerate} \) and \( \gamma_{ij} = 0 \)
The location of an IMP

\( * \) location of an IMP and a SIMP

Figure 2—The network model considered in numerical examples

for \( i=j \). The average packet lengths is assumed to be 512 bits on all ground channels. The packet length on the satellite channel is fixed and equals 1 K bit. The ground link capacities are all assumed to be \( C_g = 50 \text{ K bits/sec} \) and the satellite capacity to be \( C_s = 1500 \text{ K bits/sec} \). With this information as input to the routing optimization program, the inter-regional \( g_{ij} \) are computed and are given in Table I. Note that we assume \( g_{ij} = 1 \), for \( i, j \) in the same region.

### LINK CAPACITY ASSIGNMENTS

In the packet routing studies above we assume that the ground link capacities are given. We also assume

as given the capacity of the satellite channel. The capacity assignment problem is: for each channel \( l \) the capacity \( C_l \) must be found which minimizes the total average message delay \( T \). The problem is most difficult if the capacities must be chosen from a discrete set of options. Following Kleinrock \(^2\) we assume that the capacities of ground links and satellite channel are continuous variables and use analytic procedures involving Lagrange multipliers to obtain \( C_l \). A general discussion of our approach follows.

We assume as given the topology of the network including SIMP locations. We also assume a given demand matrix \( [Y_{ij}] \). Let \( \lambda_l \) be the traffic on link \( l \). The (continuous) capacity assignment problem can be formulated as

\[
\min_{C_l} T = \frac{1}{r} \left[ \sum_{l=1}^{L} \lambda_l T_l + \lambda_s T_s \right]
\]

subject to the constraint that the overall budget \( B \) is specified where \( B \) is

\[
B = \sum_{l=1}^{L} b_l C_l + b_s C_s
\]

and \( b_l \) and \( b_s \) are cost functions of the capacities \( C_l \) and \( C_s \), respectively.

In solving the capacity assignment problem, we need to know the total allowable budget and unit cost of ground and satellite channel capacities. We assume the unit cost of ground channel capacities are the same, so we can normalize all costs by a unit cost of ground channel capacity. In obtaining the following results, we further assume a ratio of 1 to 10 for satellite and ground channel capacity costs, i.e.,

\[
\text{1 unit cost of } \frac{1}{10} \text{unit cost of ground channel capacity} = \text{satellite channel capacity}
\]

Our program is sufficiently general so that we can assume any ratio between satellite capacity and ground capacity costs. Perhaps a more realistic ratio to use today is a 1 to 3 proportion.

Without going into the details of the capacity assignment optimization scheme, which is quite similar to the work of Kleinrock\(^2\) and which is described elsewhere,\(^1\) let us discuss a computer program we have written which combines the routing and capacity assignment algorithms in an overall procedure. The program first obtains a solution for the routing assignments including overall average delay \( T_{\text{in}} \), the routing indices \( g_{ij} \), and the channel traffic rates \( \lambda_l \) and \( \lambda_s \). Then with the \( \lambda_l \) and \( \lambda_s \) from the routing calculations and with a given total budget \( B \) and the assumed unit costs of satellite and ground channel capacities \( b_s \) and \( b_l \) as inputs to the capacity assignment routine, we obtain the capacity assignments \( C_l \) and \( C_s \) and a new figure for overall average delay \( T_e \). If \( T_e \leq T_{\text{in}} \) then the entire procedure is repeated with the previously computed values of \( C_s \), \( \{C_l\} \) and \( g_{ij} \) used as input to the routing subroutine.
The process is iterated until $T_r = T_a$ to within a specified accuracy or until the maximum time limit allowed for the optimization procedure is reached.

Using the output data from the routing subroutine run given in the previous section, the iterative cycle is carried out for the network in Figure 2. The input data consists of the original demand matrix, initial values of $C_i = 50 \text{ K bits/sec}$ and $C_s = 1500 \text{ K bits/sec}$ and an assumed total budget $B$ in cost units and $b_i = 1$ and $b_s = 0.1$. The final results for $(C_i, C_s)$ and $(\lambda_i, \lambda_s)$ after 8 iterations for $B = 1,700,000$ cost units are given in Table II, and those for $B = 950,000$ cost units after 4 iterations are given in Table III. For the total budget $B = 1,500,000$ the minimum overall average delay was $T = 0.037 \text{ sec}$ whereas for $B = 950,000$, $T = 1.435 \text{ secs.}$

We have performed a number of runs with different values of budget $B$ and the results show that as total allowable budget increases the overall average delay decreases as seen in Figure 3. Also the optimum satellite capacity decreases and the ground channel capacities increase as the total allowable budget increases. This is because we have assumed a 10 to 1 ratio between the cost of unit ground capacity to unit satellite capacity. As the total budget grows we have more money to spend on the relatively more expensive ground channel capacities. As ground channel capacities become large, packets tend to go via ground subnet since the minimum satellite propagation delay is high (0.26 secs). We can show in general that the $T$ vs $B$ curve is a hyperbola and have derived formulas for the asymptote $S$ which are given in a companion paper.\(^{17}\)

### Table II—Optimum Traffic Rates and Channel Capacities for $B = 1,700,000$ Cost Units

<table>
<thead>
<tr>
<th>Satellite Channel</th>
<th>Optimum Traffic Rates ($\lambda_i$) (packets/sec)</th>
<th>Optimum Capacities ($C_i$) (K bits/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$2.15$</td>
<td>$159.866$</td>
</tr>
<tr>
<td></td>
<td>$67.85$</td>
<td>$55.196$</td>
</tr>
<tr>
<td></td>
<td>$119.80$</td>
<td>$88.082$</td>
</tr>
<tr>
<td></td>
<td>$92.43$</td>
<td>$70.967$</td>
</tr>
<tr>
<td></td>
<td>$159.84$</td>
<td>$112.292$</td>
</tr>
<tr>
<td></td>
<td>$159.43$</td>
<td>$112.226$</td>
</tr>
<tr>
<td></td>
<td>$92.46$</td>
<td>$70.985$</td>
</tr>
<tr>
<td></td>
<td>$120.00$</td>
<td>$88.201$</td>
</tr>
<tr>
<td></td>
<td>$119.81$</td>
<td>$88.088$</td>
</tr>
<tr>
<td></td>
<td>$67.80$</td>
<td>$55.097$</td>
</tr>
<tr>
<td></td>
<td>$67.81$</td>
<td>$55.109$</td>
</tr>
<tr>
<td></td>
<td>$120.09$</td>
<td>$88.259$</td>
</tr>
<tr>
<td></td>
<td>$119.67$</td>
<td>$88.001$</td>
</tr>
<tr>
<td></td>
<td>$92.46$</td>
<td>$70.984$</td>
</tr>
<tr>
<td></td>
<td>$159.55$</td>
<td>$112.235$</td>
</tr>
<tr>
<td></td>
<td>$159.33$</td>
<td>$112.163$</td>
</tr>
<tr>
<td></td>
<td>$92.46$</td>
<td>$70.987$</td>
</tr>
<tr>
<td></td>
<td>$120.00$</td>
<td>$88.203$</td>
</tr>
<tr>
<td></td>
<td>$119.71$</td>
<td>$88.025$</td>
</tr>
<tr>
<td></td>
<td>$67.90$</td>
<td>$55.165$</td>
</tr>
</tbody>
</table>

The minimum delay = 0.0369 seconds

### Table III—Optimum Traffic Rates and Channel Capacities for $B = 950,000$ Cost Units

<table>
<thead>
<tr>
<th>Satellite Channel</th>
<th>Optimum Traffic Rates ($\lambda_i$) (packets/sec)</th>
<th>Optimum Capacities ($C_i$) (K bits/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$270.13$</td>
<td>$511.770$</td>
</tr>
<tr>
<td></td>
<td>$93.61$</td>
<td>$47.405$</td>
</tr>
<tr>
<td></td>
<td>$79.67$</td>
<td>$40.391$</td>
</tr>
<tr>
<td></td>
<td>$88.29$</td>
<td>$44.730$</td>
</tr>
<tr>
<td></td>
<td>$77.50$</td>
<td>$39.598$</td>
</tr>
<tr>
<td></td>
<td>$81.71$</td>
<td>$41.417$</td>
</tr>
<tr>
<td></td>
<td>$102.99$</td>
<td>$52.125$</td>
</tr>
<tr>
<td></td>
<td>$77.61$</td>
<td>$39.551$</td>
</tr>
<tr>
<td></td>
<td>$82.43$</td>
<td>$41.780$</td>
</tr>
<tr>
<td></td>
<td>$86.26$</td>
<td>$44.099$</td>
</tr>
<tr>
<td></td>
<td>$93.62$</td>
<td>$47.413$</td>
</tr>
<tr>
<td></td>
<td>$91.26$</td>
<td>$46.222$</td>
</tr>
<tr>
<td></td>
<td>$82.50$</td>
<td>$41.815$</td>
</tr>
<tr>
<td></td>
<td>$85.94$</td>
<td>$43.547$</td>
</tr>
<tr>
<td></td>
<td>$79.97$</td>
<td>$40.541$</td>
</tr>
<tr>
<td></td>
<td>$82.07$</td>
<td>$41.998$</td>
</tr>
<tr>
<td></td>
<td>$103.11$</td>
<td>$52.185$</td>
</tr>
<tr>
<td></td>
<td>$79.97$</td>
<td>$40.639$</td>
</tr>
<tr>
<td></td>
<td>$80.43$</td>
<td>$40.773$</td>
</tr>
<tr>
<td></td>
<td>$89.24$</td>
<td>$45.257$</td>
</tr>
<tr>
<td></td>
<td>$91.14$</td>
<td>$46.166$</td>
</tr>
</tbody>
</table>

The minimum delay = 1.435 seconds

### THE MASTER SCHEME

In this section we will describe a scheme in which retransmission traffic from a multi-access channel is sent via the ground net. This scheme will be denoted here as the MASTER (Multi-Access Satellite With Terrestrial Retransmission) plan. Under the MASTER plan, the up-link of the satellite channel is a slotted multi-access channel and the down-link is a broadcast channel. Unlike slotted ALOHA, whenever a packet collision occurs in the multi-access channel, the MASTER scheme does not use the satellite channel for retransmission of the rejected packet. Instead the rejected packet is sent via the ground net from the source SIMP directly to the destination IMP without having to first go to its regional SIMP. By sending retransmitted packets via the ground net, overall average delay can be reduced significantly especially if the traffic on the satellite channel is heavy. Moreover, the MASTER plan eliminates all possibilities of instabilities that might occur in a slotted ALOHA channel under very heavy traffic situations.\(^{17}\) Thus we see that the MASTER channel has no retransmitted packets. Since on the down-link broadcast channel each SIMP can hear its own transmissions, if a packet collision occurs, the source SIMP could reroute the rejected packet through the ground net without waiting for an acknowledgment from the destination SIMP. Moreover, after sending a packet, a SIMP can immediately send again in the next slot without waiting for the results of previous transmissions.
Now let us turn our attention to a more detailed discussion of the MASTER plan. Consider the satellite channel model depicted in Figure 4. Let $\lambda_a$ [pkts/sec] be the average input rate from SIMP $a$ to the satellite channel. The probability $q_a$ that a packet will be transmitted from SIMP $a$ in a given time slot, or equivalently the traffic rate in a given time slot is:

$$q_a = \frac{\lambda_a}{\mu_a C_s}$$

where $C_s$ [bits/sec] is satellite channel capacity and $1/\mu_a$ [bits/sec] is the packet length.

We define random sequences $X_a(k)$ and $Y_a(k)$ as

$$X_a(k) = \begin{cases} 1 & \text{if SIMP } a \text{ transmits in time slot } k \\ 0 & \text{otherwise} \end{cases}$$

$$Y_a(k) = \begin{cases} 1 & \text{if SIMP } a \text{ successfully transmits in time slot } k \\ 0 & \text{otherwise (i.e., no transmission or collision)} \end{cases}$$

Then it follows that

$$\Pr[X_a=1] = q_a$$

By assuming that the streams $X_a(k)$ from different SIMPs are independent we readily obtain

$$\Pr[Y_a=1] = q_a P_a$$

where $P_a = \Pi_{i \neq a} q_i$, with $q_i = 1 - q_i$, is the probability of success of a packet from SIMP $a$. The normalized throughput $S$ or successful transmission per slot time, of the satellite channel is thus

$$S = \sum_a \Pr[Y_a=1] = \sum_a q_a P_a$$

Note that packets from SIMPs traverse the satellite channel once and only once under this operational scheme. If we assume the satellite channel transmission errors and nodal processing delay to be small, and buffer storage to be sufficiently large, then the average delay incurred by a packet when going through the channel can be found by using response time formulation of an $M/D/1$ queueing system plus the satellite propagation delay $\tau_{min}$ of approximately .26 secs. The result of a deterministic service time system is used, since in a slotted channel all packets are of fixed length, any messages with length less than a specified full packet size are filled with blank characters to form a full size packet. Thus the average delay experienced by a packet travelling from SIMP $a$ through the satellite channel is

$$T_a = \tau_{min} + \frac{1}{\mu_a C_s} \left( 1 + \frac{\lambda_a}{2 (\mu_a C_s - \lambda_a)} \right)$$

The overall average packet delay of the satellite channel $\tau_s$ [secs/pkt] can be obtained by averaging over all $T_a$'s

$$\tau_s = \frac{1}{\lambda_s} \sum_a \lambda_a T_a$$

$$= \tau_{min} + \frac{1}{\mu_s C_s} \left( 1 + \frac{1}{2 \lambda_s} \sum_a q_a^2 \right)$$

where

$$\lambda_s = \sum_a \lambda_a$$

and

$$q_a = \frac{\lambda_a}{\mu_s C_s}$$

It can be shown that for a given total traffic rate $\lambda_s = \sum \lambda_a$ [pkts/sec] $\tau_s$ takes on its minimum when $\lambda_a = \lambda_s/M$, for all $a=1,2,\ldots,M$. We can also show that the condition $\sum q_a = 1$ is necessary to achieve the maximum throughput $S$. From these conditions we immediately see that the best throughput delay trade-off is attained for $q_a = 1/M$, $a=1,2,\ldots,M$.

This condition, although the optimum operating point, is difficult to attain, since the traffic rates from SIMPs are most likely not equal. It is difficult to study
the trade-off for cases with a large number of heterogeneous SIMPs. We can obtain some understanding of how $S$ varies by investigating the case of $M=2$. For this case, the throughput rate of the satellite channel is

$$S = q_1(1-q_1) + q_2(1-q_2) = q_1 + q_2 - 2q_1q_2$$  \hspace{1cm} (8)

A plot of the throughput $S$ versus the traffic rates $q_1$ and $q_2$ is given in Figure 5. In the throughput as shown in Figure 5, the surface takes the form of a saddle with the valley located on the plane $q_1 = q_2$ which includes the $S$-axis and is at a 45° angle from the planes $q_1 = 0$ and $q_2 = 0$ (the slashed, diagonal plane shown in Figure 5 and ridge located on the plane $q_1 + q_2 = 1$ which is perpendicular to the plane $(q_1 = q_2)$ and passes by upper-corner end points $(1,0,1)$ and $(0,1,1)$ of the throughput surface). Note that the condition

$$q_1 = q_1 + q_2$$  \hspace{1cm} (9)

corresponds to a straight line of slope $-1$ on $q_1 - q_2$ plane. As stated above if

$$q_1 + q_2 = q_1 = q_2$$  \hspace{1cm} (10)

we obtain the maximum boundary of the throughput $S$. The ridge of the saddle surface of the throughput $S$ is thus the maximum boundary. For this simple case, we can further prove that, given the condition in Equation (9), $S$ attains its minimum for

$$q_1 = q_2$$  \hspace{1cm} (11)

which is the line with slope $1$ on $q_1 - q_2$ plane (the line at 45° degree angle with $q_1$-axis and $q_2$-axis). The valley of the saddle surface of the throughput is thus the minimum boundary. Also, the overall average packet delay, $\tau_w$, of the satellite channel for $m=2$ case is

$$\tau_w = \tau_{\text{min}} + \frac{1}{2CN} \left( \frac{q_1^2}{1-q_1} + \frac{q_2^2}{1-q_2} \right) + \frac{1}{\mu Cn}$$  \hspace{1cm} (12)

Again, the delay $\tau_w$ is symmetrical with respect to $q_1$ and $q_2$. In Figure 6 we show the level curves of the delay surface in the $(q_1,q_2)$ plane. Given the condition in Equation (9) the delay $\tau_w$ reaches its minimum for $q_1 = q_2$. The minima are shown by a dashed line intersecting the level curves.

The above results are preliminary. We have not computed numerical examples for the MASTER scheme. These will be given in a companion paper by the authors.\(^7\)

CONCLUSIONS

In this paper we have presented some of the important design issues for mixed media packet switching networks. Satellite packet switching has considerable promise for low cost, high bandwidth data communications. However there is inherent high delay in satellite links which do not appear in ground links. Therefore a mix of the two communications media seems to offer the best of both worlds. In this paper we have examined a number of trade offs which offer guidelines for the design and optimum utilization of mixed media networks.

We have introduced a new communications scheme called MASTER (Multi-Access Satellite with TErrestrial Retransmission). We believe that the MASTER scheme offers significant advantages over slotted ALOHA especially when the multi-access satellite channel is heavily loaded.

We have not explored the possibility of sending network control information along the ground and using

![Figure 5: Throughput surface given by equation (8)](image)

![Figure 6: Level curves of delay surface given by equation (12)](image)
the satellite for bulk data transmission. However the extension is logical and not difficult to analyze. Another logical extension of the MASTER scheme is to use the satellite channel on a reservation basis, such as suggested by Crowther, et al.,10 Roberts11 and Binder,12 but use the ground channel to set up reservations. We plan to explore this idea in a subsequent paper.

REFERENCES


