JB 92-12

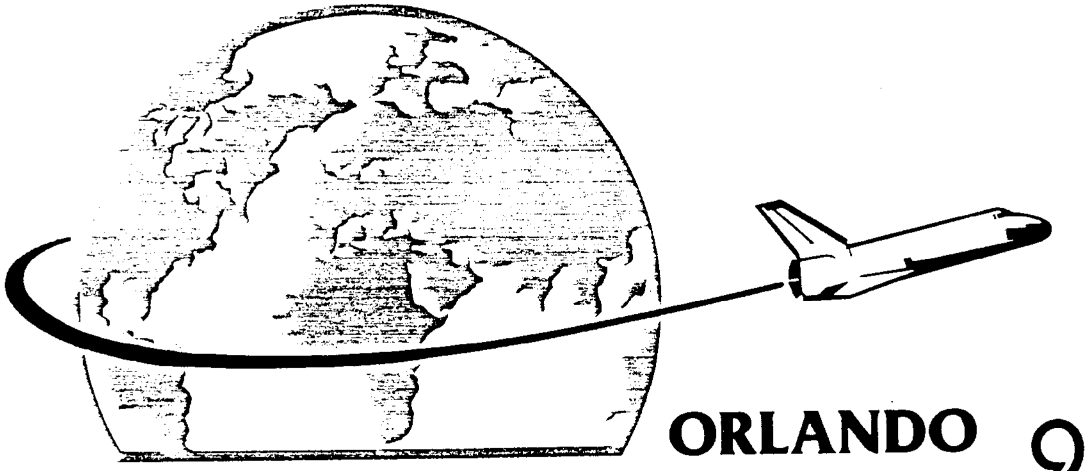
“Optimal Routing in Mixed Media Networks with Integrated Voice and Data Traffic”

(with Shihwei Chen)

Globecom' 92
Orlando, Florida
December 6-9, 1992

92-12

DECEMBER 6-9, 1992



ORLANDO

GLOBECOM '92

COMMUNICATION FOR GLOBAL USERS

IEEE GLOBAL TELECOMMUNICATIONS
CONFERENCE • ORLANDO • FLORIDA

CONFERENCE RECORD

VOLUME 1 OF 3

ICC
GLOBECOM
Sponsored by the
IEEE Communications Society
and the Orlando IEEE Section

IEEE COMMUNICATIONS SOCIETY

IEEE

Optimal Routing in Mixed Media Networks with Integrated Voice and Data Traffic [*]

Shihwei Chen and John S. Baras [†]
Electrical Engineering Department and
Systems Research Center
University of Maryland
College Park, MD 20742

Abstract

In this paper, we consider a large mixed-media network which consists of a low-delay terrestrial sub-network and a high-bandwidth satellite sub-network. Both voice and data traffic are transmitted and routed through the same network. We show how to route both traffic via ground and/or satellite links by means of static, deterministic procedures. Two common voice/data integrated protocols such as fixed boundary and movable boundary schemes for the satellite channel are investigated, and the performance of both schemes is evaluated. The optimal splitting ratios for voice and data at the SIMPs (Satellite Interface Message Processors) are found using a powerful numerical optimization package (FSQP).

1. Introduction

A mixed-media [1] network is an integrated, heterogeneous network which consists of several sub-networks. The transmission media of these sub-networks are different. For example, these sub-networks could be coarsely classified as a terrestrial network or a radio network. A terrestrial network could be a telephone network or a computer network. A radio network could comprise a satellite network or a cellular mobile network, etc.

The study of mixed-media networks is important because the overall network efficiency can be considerably increased by using all available resources and media. In some situations, the combined use of all media may provide connectivity, whereas use of a single medium may not. The main problems addressed here are the design of multi-access schemes and routing algorithms.

Huynh et al. [3] presented an early approach to the optimal design of routing and capacity assignment in mixed-media packet-switched network consisting of a ground subnet and a satellite subnet. In their algorithm, they assumed linear cost-capacity functions for both terrestrial and satellite links and a fixed-split routing policy in their terrestrial sub-network. The first assumption makes their capacity assignment problem mathematically tractable and thus a closed-form solution was obtained using analytic procedures involving Lagrange multipliers. The second assumption eliminates the routing problem within the terrestrial sub-network and reduces the flow assignment problem to one of determining the optimal amount of traffic which goes through satellite links for each pair of source/destination nodes. However, these authors considered only one uniform transmission and switching mode, i.e., only packet-switching method for data transmission among source/destination pairs.

To consider voice/data integrated systems, we must modify the objective cost function to include the performance measure for voice traffic. For example, Gerla and Pazos-Rangel [2] considered the bandwidth allocation and routing problem in ISDNs, using a linear combination of the blocking probability and packet delay as their objective function. They formulated the problem as a constraint nonlinear programming problem, which has a special structure to be exploited and solved by Frank-Wolfe's steepest descent algorithm. The previous authors considered the voice/data integration system in a single domain, not in the mixed-media network domain. It is the main theme of this paper to explore optimal routing in mixed-media networks with integrated voice and data traffic.

2. Network model

A mixed-media network could comprise several sub-networks. However, to simplify the problem, we consider a communication network composed of two sub-networks: one is the ground subnet, the other the satellite subnet. The nodes are the locations of interface message processors (IMPs) linked together by landlines. There are special nodes called satellite IMPs (SIMPs) which are interface message processors between the satellites and the ground links.

Routing in a mixed-media network consists of two major portions: (1) splitting of the input traffic at SIMPs between ground and satellite subnets; (2) routing on the ground subnet. The traffic of the ground subnet consists of voice traffic

---

[*] Supported in part by grant from NASA NAGW-2777 through the CCDS program.

[†] Martin Marietta Chair in Systems Engineering.

[1] This term should be made distinctive from multimedia which means multiple traffic of voice, data, video, etc.

and data traffic. Our design problem can be stated as follows. Given a network topology, a set of input traffic rates, a ground routing procedure, and link capacities, we want to minimize a weighted sum of the average delays of data traffic in both the ground subnet and satellite subnet and the blocking probabilities of voice traffic in both the ground subnet and satellite subnet over the set of traffic splitting factors of data and voice respectively.

Let g_{ij} (s_{ij}) be the splitting factor of data traffic which specifies the fraction of data (voice) traffic, originating at node i and destined for node j, going through the ground sub-network. In the following, we will derive the overall objective function.

2.1. Data delay on the ground

Suppose that we are given a data sub-network in the ground consisting of N nodes linked by L ground links of capacities C_{dl} (bits/sec), $l = 1, 2, ..., L$, in a specified topology. The network is partitioned into M regions, each having a SIMP. These M SIMPs are linked via a satellite channel of capacity C_s (bits/sec). A traffic rate matrix $[r_{ij}]$ specifies, in packets/sec, the average rates of messages flowing between all possible IMP pairs i and j, where $i, j = 1, 2, ..., N$.

If we make the following (Kleinrock independence) assumptions: Poisson arrivals at nodes, exponential distribution of packet length, independent arrival processes at different nodes, independent service times at successive nodes, then we have the expression for average data delay in the ground [1],

$$D_g = \frac{1}{\gamma}\sum_{l=1}^{L}\lambda_{dl}T_l \qquad T_l = \frac{1}{\mu_d C_{dl} - \lambda_{dl}} \tag{1}$$

where T_l is the delay on link l; and $\gamma = \sum_{i,j=1}^{N}\gamma_{ij}$ = the total data traffic rate in the data sub-network; and λ_{dl} = the traffic rate on link l; and $\frac{1}{\mu_d}$ = the average length of a packet.

2.2. Voice blocking on the ground

Suppose that we are given a telephone sub-network on the ground which may use the same transmission links and switching facilities as the data sub-network on the ground. This voice sub-network has N_1 nodes linked by L_1 trunks (links) of capacities $C_{vl}, l = 1, 2, ..., L_1$ in a specified topology. The capacity C_{vl} of link i can be divided into N_{vl} channels. A traffic rate matrix $[\Gamma_{ij}]$ specifies, in calls/min, the average rates of call requests between all possible IMP pairs i and j, where $i, j = 1, 2, ..., N_1$. We can model each trunk by an $M/M/N_{vl}/N_{vl}$ system and the average blocking probability P_b is [1]

$$P_b = \frac{1}{\Gamma}\sum_{l=1}^{L_1}\lambda_{vl}P_l \qquad P_l = \frac{(\lambda_{vl}/\mu_v)^{N_{vl}}/N_{vl}!}{\sum_{n=0}^{N_{vl}}(\lambda_{vl}/\mu_v)^n/n!} \tag{2}$$

where P_l is the blocking probability of trunk l; and $\Gamma = \sum_{i,j}^{N_1}\Gamma_{ij}$ = the total voice traffic in the voice sub-network; and λ_{vl} = the traffic rate on link i; and $\frac{1}{\mu_v}$ = the average holding time of a phone call.

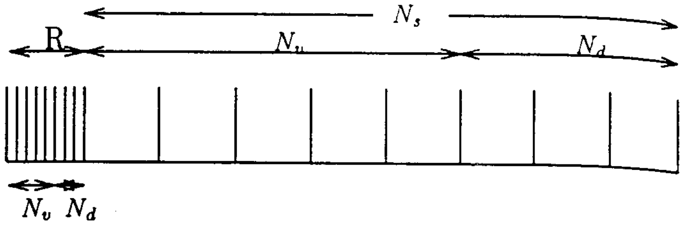


Figure 1: The satellite channel assignment

2.3. Satellite channel

To integrate voice and data traffic in the satellite channel, we can have two strategies: a fixed boundary strategy or a movable boundary strategy. In the fixed boundary strategy, the data packets are not allowed to use the voice channels even if some of them are idle. In the movable boundary strategy, the data packets can occupy any of the voice channels not currently in use. However, an arriving call has higher priority to preempt the data packets serviced in the voice channels.

We make the following assumptions in our model: (1): The SIMPs collectively generate Poisson data traffic at rate Λ_d packets/sec and Poisson voice traffic Λ_v calls/min (excluding the retransmission due to collision). The overall transmission rate (the original rate plus retransmission rate) of data and voice traffic into the satellite channel are denoted as Λ'_d and Λ'_v respectively. (2): The data packets are of fixed length. The voice call duration is exponentially distributed with mean $1/\mu_v$ seconds. (3): Channels are slotted. Let T denote the slot length which equals the transmission time of a packet and S be the round-trip delay of the channel measured in slots. (4): The retransmission delay for a request or a random access data packet is uniformly distributed between 0 and K slots.

2.3.1 Data delay in the satellite channel:

There are R reservation channels and N_s message channels which are further divided as N_v voice channels and N_d data channels. The word "message" here refers to either voice calls or data packets. All channels are slotted. The length of a slot time is equal to the transmission time of a data packet. A time slot is further divided into n minislots for message reservation and the length of each minislots (T/n) is equal to the transmission of a request packet. There are N_s (nR) minislots in the R reservation channels. Among the N_s minislots, the first N_v are used for voice requests and the others ($N_s - N_v$) are used for data requests. See Fig. 1.

Define p_{suc} to be the probability that a data packet will be successful on a data channel. Then p_{suc} is $\frac{\Lambda'_d}{m}Te^{-\frac{\Lambda'_d}{m}T}$ (see [5] page 430). Here we assume the data packets can go to any of the m data channels and the service at each channel is independent of other channels' services. The throughput of m (could be N_d in the fixed boundary scheme or a variable in the movable boundary scheme) data channels system η_{RAD} is $m \times p_{suc}$. The average delay under random access data (RAD) protocol is [6]:

$$D_{RAD}(m) = [1.5 + S + (e^{\frac{\Lambda'_d}{m}T} - 1)t_{rx}]T \qquad (3)$$

where $e^{\frac{\Lambda'_d}{m}T} - 1$ is the average number of retransmissions required for the data packet; and $t_{rx} = 1.5 + S + \frac{K-1}{2}$ is the average retransmission time measured in slots; and $\Lambda_d = \Lambda'_d e^{-\Lambda'_d T}$ [5]. According to the same reference, the queueing delay at each SIMP is neglected.

2.3.2 Voice blocking in the satellite channel: For voice channel, this is an $M/G/N_v/N_v$ system, and the average blocking probability P_s is given by

$$P_s = \frac{\Lambda_v}{\Gamma} P_v \quad P_v = \frac{(\Lambda_v/t_v)^{N_v}/N_v!}{\sum_{k=0}^{N_v} (\Lambda_v/t_v)^k/k!} \qquad (4)$$

where P_v is the blocking probability of the satellite channel; and $\Lambda_v = \sum_{\sigma=1}^{M} \Lambda_{v\sigma}$ is the overall voice traffic rate into the satellite channel; and $\Lambda_{v\sigma}$ =the voice call arrival rate from SIMP σ into the satellite channel; and $1/t_v$ is the average call duration plus the round-trip delay ST plus the call request and set-up time.

Since the typical call duration is much longer than the round-trip delay and call set-up time, we can further simplify the system to an $M/M/N_v/N_v$ queue and the probability that a system with N_v channels has n active voice calls is given by

$$\pi_v(n) = \frac{(\frac{\Lambda_v}{\mu_v})^n/n!}{\sum_{k=0}^{N_v} (\frac{\Lambda_v}{\mu_v})^k/k!} \qquad (5)$$

2.3.3 Fixed boundary scheme: Under the fixed boundary strategy, the data packets are not allowed to use the voice channel. The transmissions of voice calls and data packets do not affect each other. Thus, the performance analysis are the same as in sections 2.3.1 and 2.3.2

2.3.4 Movable boundary integrated protocol: The data packets can use the idle voice channels in this strategy. To simplify the calculations, we can assume the data queues reach their stationary state when k, $0 \le k \le N_v$, voice calls are active. This is reasonable because the average call duration is much larger than call request and set-up time which includes propagation delay and random retransmission delay in the satellite channel. The average packet delay is

$$D_{MB} = \sum_{k=0}^{N_v} \pi_v(k) d_{data}(N_v - k) \qquad (6)$$

where d_{data} is obtained from r Equation (3); N_v could be as large as N_s which is a constant. In this case, all the channels are used by the voice traffic and data traffic uses the channels not occupied.

2.3.5 Average delay in the satellite channel: The overall average delay D_s in the satellite channel is

$$D_s = \frac{\Lambda_d}{\gamma} D_d \qquad (7)$$

where D_d can be D_{RAD}, D_{MB}, or other analytical expressions of other data access schemes; and $\Lambda_d = \sum_{\sigma=1}^{M} \Lambda_{d\sigma}$; and $\Lambda_{d\sigma}$ = the data arrival rate from SIMP σ into the satellite channel.

3. Problem formulation

The overall objective function which we want to minimize is thus given as

$$f = A \times D_g + B \times P_b + C \times D_s + D \times P_s$$

where A, B, C, D are the weighting factors which can be adjusted according to different network topologies. For example, $A = 0$ ($B = 0$) means that the ground sub-network doesn't have any data (voice) traffic or the delay (blocking probability) in the ground sub-network is not an important factor in our consideration. Therefore, our formulation is more general than previous works. Huynh et al. [3] considered only data traffic in packet-switched networks, which is a special case of ours by setting $A = C = 1, B = D = 0$.

To summarize, the design of the routing problem can be formulated as following:

$$\text{minimize} \quad f(\lambda_{dl}, \lambda_{vl}, \Lambda_d, \Lambda_v)$$

$$\text{subject to} \quad 0 \le g_{ij} \le 1, \quad 0 \le s_{ij} \le 1$$

3.1. Numerical examples

The optimization problem can be solved by different algorithms. Among those are primal-dual methods, and Sequential Quadratic Programming (SQP). We have used the FSQP (Feasible SQP) subroutines developed by J. Zhou and A. Tits at the University of Maryland.

[Example 1:] First, we consider a network with data traffic only. This example is taken from [3]. This network has eight nodes (IMPs) and 20 links as shown in Fig. 2. In this network, there are two regions consisting of nodes {1, 2, 3, 4} and {5, 6, 7, 8} respectively; and the regional SIMPs are located at nodes 1 and 7, which are also IMPs. The traffic demand matrix is assumed to be uniform with $\gamma_{ij} = 12$ (packets/sec) for all $i \neq j$ and $\gamma_{ii} = 0$ for all $i = 1, 2, ..., 8$, and the average packet length is assumed to be 512 bits on all ground channels. The packet length on the satellite channel is fixed and equals 1 kbits. The ground link capacities (C_l) are all assumed to be 50 kbits/sec (5×10^4 bits/sec), and the satellite capacity to be $C_s = 1.5 \times 10^6$ bits/sec.

The routing indexes g_{ij} are given in Table 1 after running the FSQP. The overall objective function is the combination of average delays in ground links and satellite link. Due to the symmetry of the given network topology and the uniformity of the traffic requirements among all SD pairs,

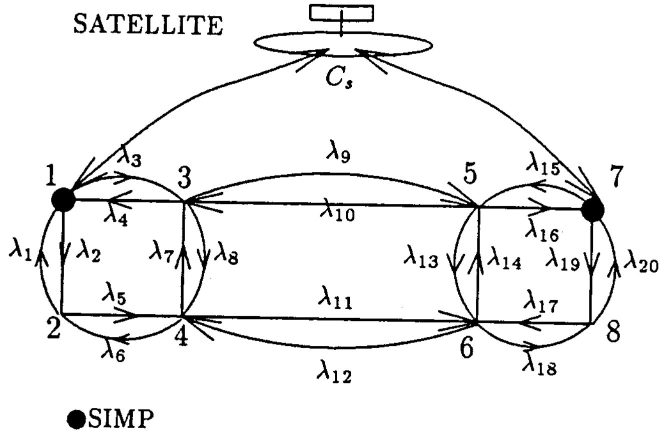


Figure 2: A mixed-media network

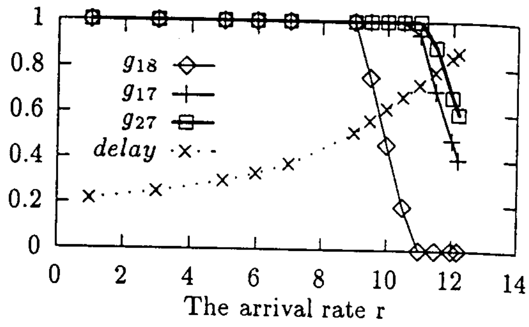


Figure 3: The effects of increasing arrival rate r_{ij}

Table 1: Splitting ratio g_{ij} for a data network

		destination			
		5	6	7	8
	1	1	1	0.482	0
	2	1	1	0	0.677
origin	3	1	1	1	1
	4	1	1	1	1
		objective=0.0842 sec.			

Table 2: Splitting ratio s_{ij} for a voice network

		destination			
		5	6	7	8
	1	0	0	0	0
	2	0.559	1	0	0
origin	3	1	1	0.177	0.322
	4	1	1	0	1
		objective=0.98			

we will expect the results to demonstrate symmetry, too. However, the results in Huynh's original paper didn't show any symmetry. This is why we examined again the system. Consider the delay as a function of requirement traffic matrix γ_{ij} only. We have the following observations.

[Observation 1:] The delay is a convex function of γ_{ij}. When the rate is small, all the traffic goes through ground links, since the delay of the ground sub-network is small. As the rate increases, the ground links becomes loaded and eventually saturated when the arrival rates approach link capacities, more and more traffic will go through satellite links. The relation between γ_{ij} and g_{ij} is shown in Fig. 3. The result shows that the SD pairs with more hops will deviate their traffic to satellite first (e.g. SD pair 1, 8) and then SD pairs with fewer number of hops (SD pair 1,7, then pair 2,7), etc. This is the so-called "farthest end routing" in [4]. The last to sent their data through satellite are the nodes one-hop away (SD pair 3,5 and 4,6). This observation is also true for other network topologies.

[Observation 2:] The symmetry of our example can be represented in equivalence classes. For example, nodes {1, 2, 7, 8} and {3, 4, 5, 6} are equivalence classes. There the traffic from IMPs 8,1 are the same as that from IMPs 7,2, and $g_{1,8}$ will be the same as $g_{8,1}$. Symmetric results like this can be used to check the correctness of the programming. Of course, just symmetry is not sufficient to guarantee the correctness of the results and this observation is only applied to the networks with symmetric topologies and uniform traffic requirement rates among nodes.

[Example 2:] In this example, we consider a network with voice traffic only, the objective function is thus the summation of blocking probabilities of ground links and satellite links. The network is the same as that in Example 1. However, the capacities are converted into channels for voice transmission. The traffic requirement matrix is again assumed to be uniformly 1 calls/min, and the average call duration is 4 minutes per call, and the satellite has 50 channels and the ground links have 5 channels per link. The voice splitting ratios s_{ij} are obtained in Table 2.

We note that the difference of voice traffic and data traffic is that even at low traffic rates some traffic, except that of inter-region nodes, goes to satellite channel. This is because satellite has much larger capacity than any of the ground links and delay of the satellite is not a big factor to consider. The ground links inside a region still have intra-region calls to transmit, though their traffic to another region will go through satellite link. In contrast, the traffic of neighboring inter-region nodes only uses the ground links.

[Example 3:] In this example, we consider a ground subnetwork capable of transmitting and switching both voice and data traffic through the same IMPs and transmission links. Using the same network and the same data as in

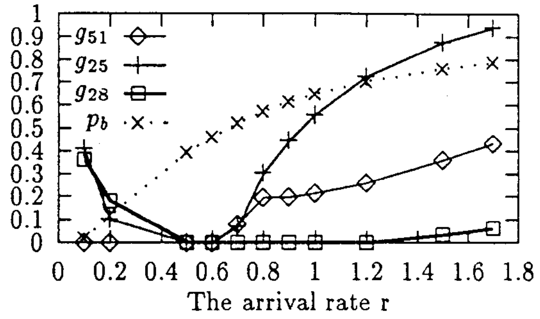


Figure 4: The effects of increasing arrival rate r_{ij}

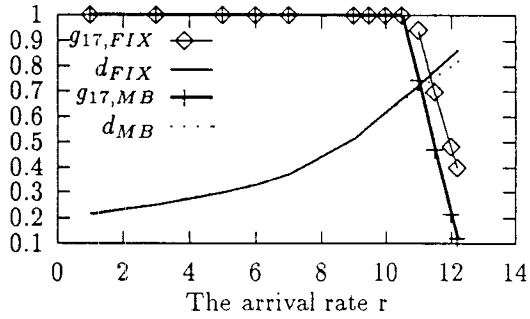


Figure 5: The effects of movable boundary scheme

Table 3: Splitting ratio g_{ij} for the data sub-network

		destination			
		5	6	7	8
origin	1	1	1	0.165	0
	2	1	1	0	0.487
	3	1	1	1	1
	4	1	1	1	1

the previous examples. For the fixed boundary scheme, we would expect the splitting ratios of voice and data will be the same as those of Example 2 and 1 respectively, since the traffic types are independent of each other. For movable boundary, we assume the following parameters of the system: the number of voice channels in the satellite is 20 and data is 10, ground capacity in each link is 5 channels/link. The traffic requirement for data is uniformly 6 packet/sec and for data is 1 call/min. The delay expression used for satellite channel is D_{RAD}. The overall objective function is the (unweighted) summation of delays and blocking probabilities in the ground links and satellite channel. We have the following results as in Table. 3 and Table 4.

We note the results are quite similar to the previous examples. From the illustration of Fig. 5, we confirm that the delay of the movable boundary scheme is lower than that of fixed boundary scheme when the arrival rate is large enough so that some traffic goes to satellite channel. We also find that more traffic will go to satellite channel under the movable boundary scheme. Therefore, the moving boundary scheme can increase the utilization.

Table 4: Splitting ratio s_{ij} for the voice sub-network

		destination			
		5	6	7	8
origin	1	0		0	0
	2	0.257	1	0	0
	3	1	1	0.176	0
	4	1	1	0	1
		objective=0.191			

4. Conclusions

We have obtained the optimal routing ratios between the terrestrial and the satellite networks for both voice and data traffic. In satellite channel, demand assignment and random access schemes are used for data access; for voice traffic, reservation must be used before paths are established. Their performance is evaluated. Our methodology allows us to treat a large network.

Acknowledgment: We would like to thank Drs. J. Zhou and A. Tits for offering the powerful FSQP optimization package.

References

[1] D. Bertsekas and R. Galleger. *Data Networks.* Prentice-Hall, 1987.

[2] M. Gerla and R. Pazos-Rangel. "Bandwidth allocation and routing in ISDN's". *IEEE Commun. Mag.*, vol. 22(no. 2):pp. 16–26, Feb. 1984.

[3] D. Huynh, H. Kobayashi, and F. Kuo. "Optimal design of mixed-media packet-switching networks: Routing and capacity assignment". *IEEE Trans. Commun.*, vol. COM-25(no. 1):pp. 156–187, Jan. 1977.

[4] Y. Morihiro, S. Okasaka, and H. Nakashima. "A dynamic channel assignment and routing satellite and digital networks-DYANET". In *Proc. GLOBECOM'87*, pages 10.1.1–10.1.5, 1987.

[5] M. Schwartz. *Telecommunication networks: Protocols, modeling and analysis.* Addison Wesley, 1987.

[6] C. Wu and V. Li. "Integrated voice and data protocols for satellite channels". In *Proc. Mobile Satellite Conference, Jet Propulsion Lab.*, pages 413–422, May 1988.