INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, VOL. 12, 147–155 (1994)

AN ADVANCED SATELLITE COMMUNICATION SYSTEM WITH ON-BOARD FAST PACKET SWITCHING CAPABILITIES

ENRICO DEL RE AND ROMANO FANTACCI
Dipartimento di Ingegneria Elettronica, Università di Firenze, Firenze, Italy

SUMMARY

This paper is focused on a satellite communication system having on-board fast packet switching capabilities. Different alternatives for the up-link access technique and for the architecture of the on-board switching fabric are considered. In particular an efficient time division multiple access technique with slots assigned on demand and a novel switching approach are proposed. Performances in terms of mean access delay and mean on-board switching delay are derived by analytical approaches and computer simulations.

KEY WORDS Satellite communications Switching systems Multiple access schemes Queueing theory

INTRODUCTION

This paper deals with a network-oriented satellite communication system in which the satellite operates as a remote node. The satellite uses multiple spot beams and has baseband fast packet switching capabilities.[1–5]

Performance of the satellite communication system can be derived in terms of end-to-end delay which is formed (apart the propagation delay) by the sum of three contributions.

A first contribution is due to the access delay of the protocol used by the earth-stations in the up-links. Different alternatives have been considered for such an access protocol, namely time division multiple access (TDMA) which provides a fixed assignment of the channel capacity to the stations of the same spot and a TDMA protocol with slots dynamically assigned on demand (DTDMA).[1,2]

A second contribution is due to the on-board switching delay introduced when two or more packets arriving simultaneously on different up-links (satellite spot beams) have the same down-link destination. One of these contending packets attains switching, whereas queueing is required for the others to wait for a later route.[3–5]

Depending on the speed of the switch fabric and its architecture, different approaches for providing the queueing necessary for contending packets are possible.[3–5] This paper is focused on the input queueing that requires hardware operating at a slower speed. In particular, with the aim to decrease the on-board switching delay, a novel input queueing approach is proposed and studied.

Finally, the third contribution is due to the transmission delay in the down-links. The asynchronous TDMA protocol (ATDMA) is assumed in transmission on to the down-links.[4] We are motivated in this choice by the statistical multiplexing of the packets performed in the on-board queue. In performing our analysis, evaluation of the transmission delay in the down-links has been included in the analysis of the on-board switching delay.

SATELLITE NETWORK MODEL

The satellite communication network considered in this paper (Figure 1) can support mixed real-time and jitter-tolerant traffic of data, voice and video services. Information transmission is organized in packets formed by a fixed number of bits. Unlike traditional data networks, in a multimedia environment it is critical to guarantee delay and loss performance requirements adequate for all applications supported: from delay-sensitive applications such as voice, to loss-sensitive applications such as image transfer. Suitable traffic control algorithms will be required to provide these performance guarantees while ensuring that sufficient bandwidth is allocated

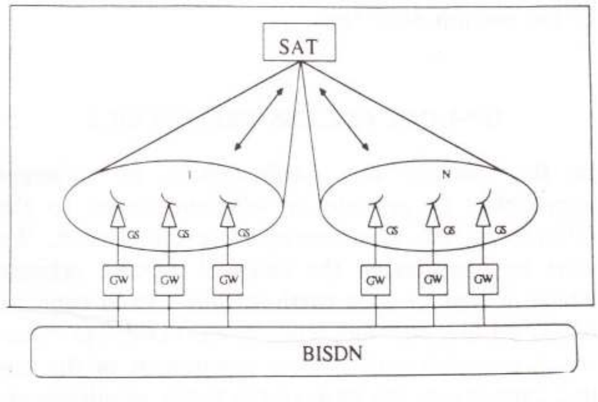

Figure 1. General scheme of the multibeam fast packet switching satellite network: GS = earth-stations; GW = Gateways; SAT = satellite

CCC 0737–2884/94/020147–09

Received September 1992
Revised November 1993

for each application type according to its specific traffic flow requirements. This paper focuses mainly on jitter-tolerant traffic for which our network offers potential gains in bandwidth efficiency by buffering and statistically multiplexing bursty traffic.

In performing our analysis we have assumed that

1. The entire coverage area is divided into N geographically disjoint spot beam zones which make frequency reuse possible in the up-links.[1,2]
2. The satellite has on-board baseband processing and switching capabilities.
3. Buffers are provided on board for messages waiting for transmission in the down-links.
4. There are N receivers and N transponders on the satellite and at most one transponder can serve the same zone at the same time.
5. Each received packet on board the satellite can have as destination any of the N possible zones with equal probability.
6. The service provided by each transponder has a frame structure, which is filled by transmitting packets (if any) from the corresponding buffer in a first-in-first-out (FIFO) discipline.

The overall end-to-end delay for the satellite network under consideration results to be

$$T_t = T_U + T_B + 2t_d \tag{1}$$

where T_U denotes the access time which elapses from the instant of arrival of the packet at the earth-station to the instant of its transmission completion in the up-link channel; T_B denotes the mean switching delay which elapses from the instant of arrival of the packet on board the satellite up to the instant of its transmission completion in the down-link channel. Note that T_B is strongly influenced by the architecture of the on-board FPS fabric. Finally, t_d denotes the earth/satellite and satellite/earth delay.

The parameter T_U will be evaluated in the next section and the evaluation of T_B will be the subject of the section after that.

UP-LINK ACCESS PROTOCOLS

In this section the performances of different approaches to provide an efficient access to the satellite by the earth-stations are considered. We start by considering the classical TDMA scheme which assigns to each earth-station a fixed capacity of the up-link channel. Then the DTDMA protocol which provides a dynamical assignment of the up-link capacity on the basis of the traffic requirements at the stations will be discussed. An optimization of the DTDMA protocol will be also performed to reduce the access delay. A performance comparison between the TDMA and DTDMA protocols will be given in term of mean access delay.

TDMA Protocol

In the TDMA protocol, time is divided into frames of fixed length. Each frame contains N_s slots with duration τ, as shown in Figure 2. At each terrestrial station messages arrive according to a Poisson process with average rate λ messages per second (mess/s). Each message is formed by a number of a fixed length packets which is geometrically distributed with mean value $\bar{m}$ equal to 5 (packets) and mean square value $\bar{m}^2$ equal to 45 (packets^2).

The theoretical analysis will be done at message departure times as is commonly used for the $M/G/1$ queueing system.[6,7] Nevertheless, in our case we must account for the special kind of service. Therefore, the usual analysis must be modified to account for the fact that the first packet of a message arriving at an empty system may not receive immediate service.[2,7] It can be required that such a packet must wait for the next slot. This additional waiting time does not arise for messages arriving at a busy system. The time diagrams for arrivals to a busy and empty system are shown in Figure 3. The service time of a message starts when that message reaches the head of the queue. The service time ends when the last packet of a message has complete transmission. In the case of Figure 3(a), the message labelled as C_n arrives at a busy system, and its service time is equal to a frame (i.e. $T_f = N_s\tau$). In the case of Figure 3(b), the message C_n arrives at an idle system; therefore it receives a different service which lasts a fractional frame followed by integral frames. It follows that the total delay in transmitting a message from an arbitrary earth-station to its destination can be considered to have three components:

(a) the message transmission time
(b) the queueing delay in the buffer of the station
(c) the slot synchronization delay.

The mean value of the slot synchronization delay for Poisson arrivals can be assumed equal to half the frame time, i.e. $N_s\tau/2$.[8]

To evaluate the mean value of the queueing delay, each station can be modelled as an independent $M/G/1$ queueing system,[4,7,8] with a mean message arrival rate per second λ and with the number of packets per message geometrically distributed. Hence, the final result for the mean access delay (normalized with respect to τ), is

$$T_U = 1 + \overline{m}N_s - \frac{N_s}{2} + \frac{N_s^2 \lambda \tau \overline{m^2}}{2(1 - N_s\lambda\tau\overline{m})} \tag{2}$$

Figure 4 shows the parameter T_U as a function of $\lambda\tau$ for different values of N_s. In the same Figure

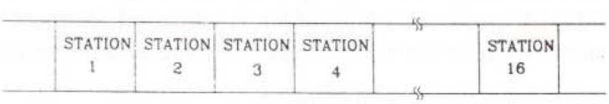

Figure 2. TDMA frame (N_s = 16)

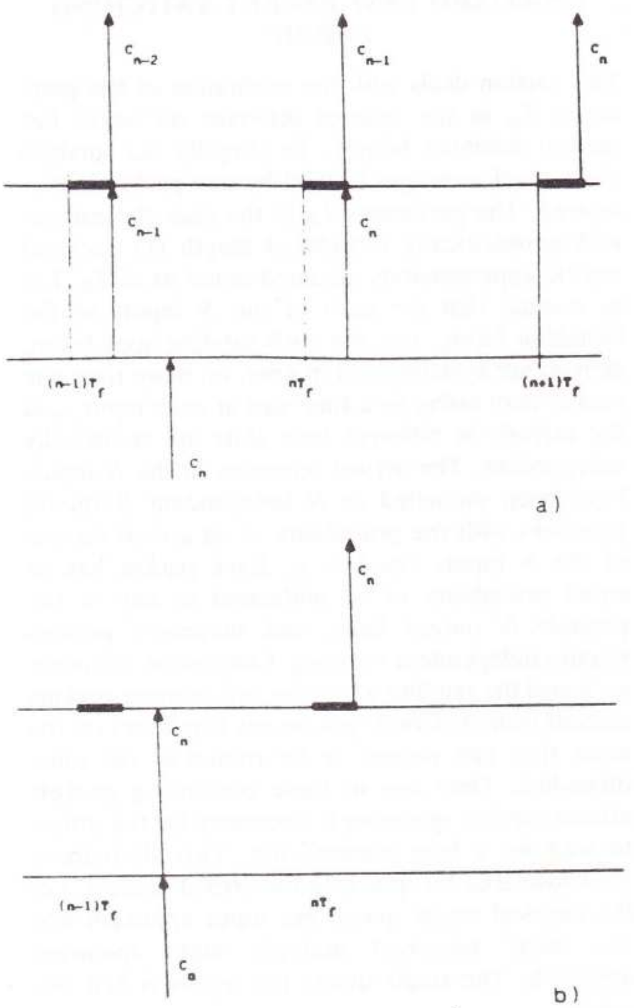

Figure 3. Time diagrams for arrivals to busy and empty systems: (a) busy system; (b) empty system

the simulation results are also reported to validate the theoretical analysis.

TDMA protocol with slots assigned on demand

The TDMA protocol discussed in the previous section permits allocation of the channel capacity to the users, independently of their activity, by assigning fixed predetermined time slots to each user. This protocol may be not a suitable choice when the traffic intensity is low or not uniform for all the stations. In this case it could be more convenient to use the DTDMA protocol presented in this section in which slots are assigned on demand.

For the DTDMA protocol under consideration the time is divided into frames of M slots of duration τ to transmit informative packets and N_s minislots of duration τ_0, such that $N_s\tau_0 = \tau$. These minislots are used by earth-stations (one for each station) to inform the channel controller on board the satellite of the number of packets which are queued waiting for service. In our model we have considered that frames are organized in superframes of duration $2t_d$. Each superframe is formed by $SM + 1$ slots (i.e. S frames plus N_s minislots). The superframe duration

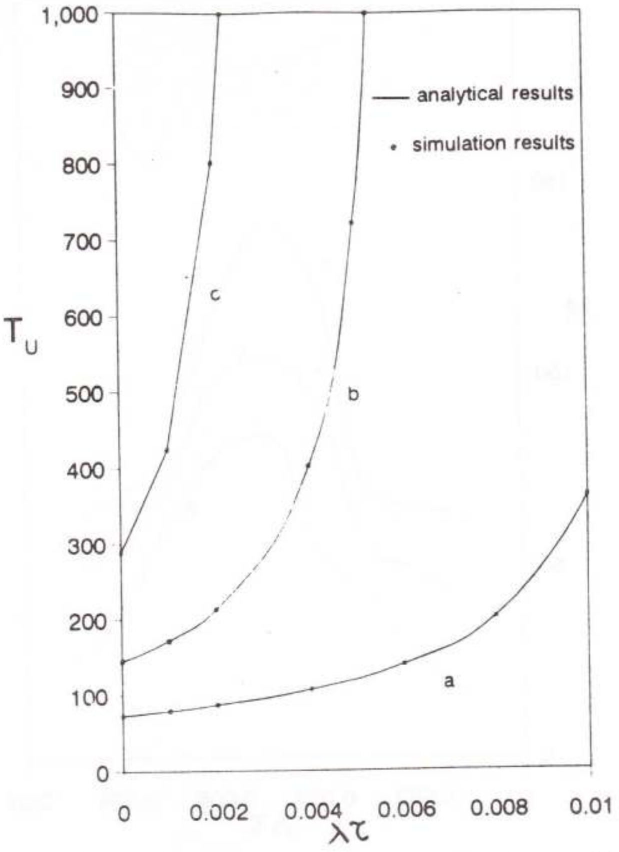

Figure 4. Up-link mean transmission delay (TDMA protocol): (a) $N_s = 16$; (b) $N_s = 32$; $N_s = 64$

has been set to $2t_d$ to achieve a good tracking of the traffic fluctuations at the earth-stations and to reduce the throughput degradation introduced by the control overhead (i.e. by the N_s minislots). The channel controller handles all the received information regarding the request for capacity assignment by the earth-stations according to the following algorithm:

1. Slots are assigned to earth-stations proportionally to the number of packets waiting for service in their buffers. However, almost one slot is assigned for each station.
2. After performing the capacity assignment algorithm the unassigned slots are equally assigned to all the stations.

   This procedure is started over again at the beginning of each superframe when an up-to-date on the status of the queue at the N_s earth-stations is received on board the satellite.

The performance of the DTDMA protocol is strongly dependent on the number of slots within each frame. To validate this assumption, Figure 5 shows optimum values of M as a function of $\lambda\tau$ for different values of N_s. Figure 6 shows the access delay achieved by using the optimum values of M. The better performance of the DTDMA protocol in comparison with the TDMA technique (Figure 4) is evident, in particular under high traffic load condition at each station.

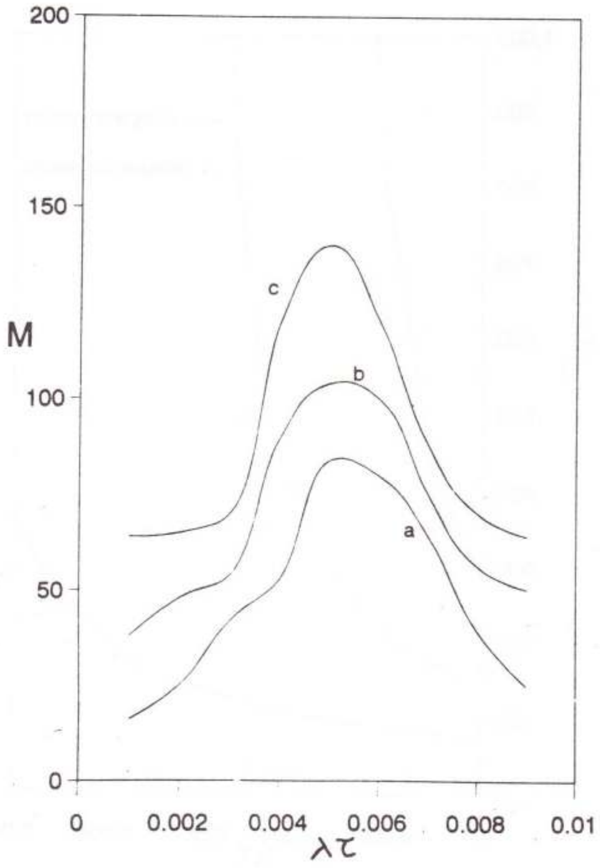

Figure 5. Optimum values of M as a function of $\lambda\tau$: (a) $N_s = 16$; (b) $N_s = 32$; (c) $N_s = 64$

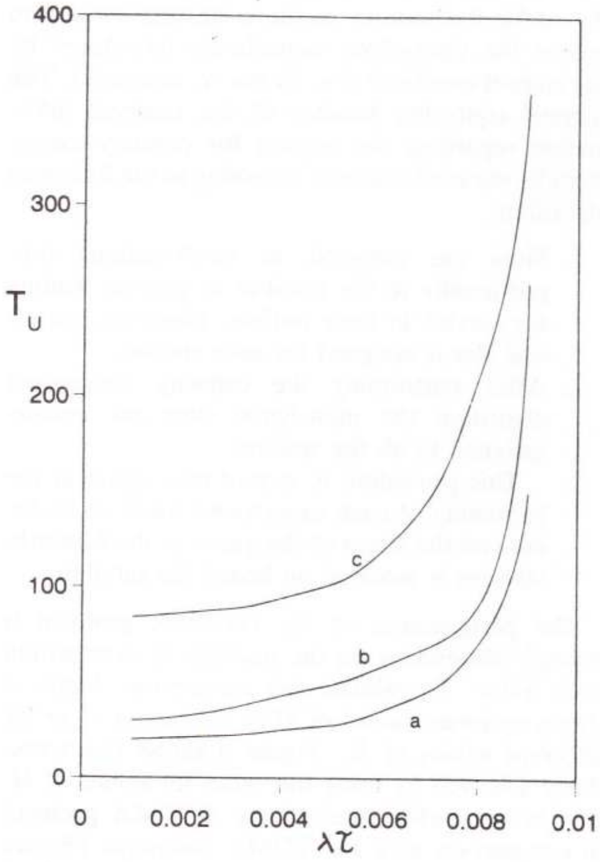

Figure 6. Up-link mean transmission delay (DTDMA protocol): (a) $N_s = 16$; (b) $N_s = 32$; (c) $N_s = 64$

ON-BOARD FAST PACKET SWITCHING FABRIC

This section deals with the evaluation of the parameter T_B in the case of different on board fast packet switching fabrics. To simplify our analysis the case of messages formed by one packet is considered. The parameter T_B in the case of messages with geometrically distributed length (in packets) can be approximately assumed equal to $\bar{m}\ T_B$. Let us assume that for each of the N inputs of the switching fabric, one for each satellite spot beam, arrivals are synchronized in time, no more than one packet may arrive in a time slot at each input, and the arrivals in different time slots are statistically independent. The arrival processes at the N inputs have been modelled as N independent Bernoulli processes with the probability of an arrival on one of the N inputs equal to p. Each packet has an equal probability to be addressed to any of the possible N output links, and successive packets require independent routeing. Congestion can occur on board the satellite whenever two or more packets arrived from different spot beams (up-links) on the same time slot require to be routed to the same down-link. Only one of these contending packets attains service; queueing is necessary for the others to wait for a later transmission. Two alternatives are considered for queueing the arrival packets, i.e. the classical single queue per input approach and the novel proposed multiple input queueing approach. The single queue per inputs is first discussed.

The single queue per input approach

The switching fabric with single queue per input considered in this section is shown in Figure 7. Whenever a cell is stored in an input buffer (IB) a routeing request packet (RRP) is broadcast over the associated bus, one for each input, to all the output controllers (arbiters in Figure 7). Collisions over the same bus are impossible because almost one cell may arrive per slot at each input port. Each RRP contains

(a) the address of the switch output (down-link) to which the packet is destined
(b) a single activity bit, to inform arbiters about the presence (logic 1) or absence (logic 0) of RRPs to be processed at their inputs.

At the beginning of each time slot, the path through each of the N address filters (AF in Figure 7) is open, initially allowing all arriving RRPs to pass through to the arbiters. The output address bits for each arriving RRP are compared bit-by-bit against the output address of all AFs, one for each possible output port. If at any time the address of an RRP differs from that of an AF, the further progress of the RRP to the arbiter is blocked. That is the output of the AF is set at logic 0 for the remainder of the

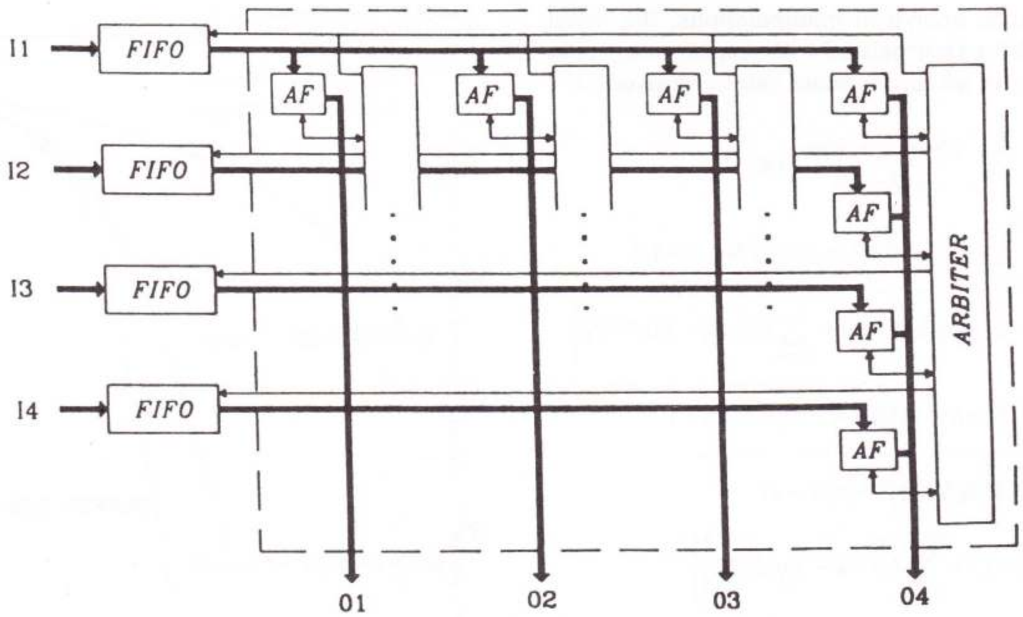

Figure 7. On-board switching fabric scheme (input queueing)

time slot. By the end of the output address, the AF will have either blocked the RRP, and hence also set its activity bit to 0, or allowed to the RRP to continue on to the arbiter. Note that even though a portion of the address bits of a blocked RRP passes through the filter, these bits are no longer processed by the arbiter, as the activity bit of that RRP has been set to 0.

Packets are routed to the appropriate output link strictly in their order of arrival: this means that the order of service is FIFO within each input queue. Therefore, the basic arbiter configuration can be realized using a simple FIFO buffer. Any new arriving RRP is stored in the FIFO buffer to form a routeing requests queue (RQ).

From the above, it follows that the switching operation is performed here in two stages. In stage one, the RRP associated with each packet is analysed, whereas in stage two the packet itself is transmitted into the output link, whenever the associated RRP reaches the head of the appropriate RQ.

Focusing our attention on a packet at the head of an input queue, i.e. the tagged packet, its service time is equal to the total time which must elapse until its RRP (the tagged RRP) leaves the appropriate RQ (the tagged RQ). Note that any RRP leaves its RQ when the transmission of the associated packet on the appropriate down-link is completed. It follows that each input queue can be modelled as a discrete $M/G/1$ queueing system.

The embedded Markov chain is used to derive T_B. The embedded points are considered as the instants of service completion. The probability generating function of the number of packets in the input queue, assuming equilibrium, is[5]

$$Q(z) = \frac{Q_0 A(z)(z-1)}{z - A(z)} \tag{3}$$

where Q_0 is the probability of having an idle queue, and $A(z)$, the probability generating function of arrivals during a service period, is defined as

$$A(z) = (1 - p + pz) \sum_{k=0}^{N-1} \frac{1 - (N - 1 - p^2 + p^2 z)^{N-k}}{(N-k)\,(N-1)^{N-k-1} p^2 (1-z)} P(k) \tag{4}$$

In (4), $P(k)$ is the probability of having k RRPs waiting for service in the tagged RQ, defined as

$$P(1) = \frac{1 - \left(1 - \frac{p}{N-1}\right)^{N} - \frac{Np}{N-1}\left(1 - \frac{p}{N-1}\right)^{N-1}}{\left(1 - \frac{p}{N-1}\right)^{N-1}} P(0) \tag{5}$$

$$P(k) = \frac{1 - \binom{N-k}{1}\frac{Np}{N-1}\left(1 - \frac{p}{N-1}\right)^{N-k-1}}{\left(1 - \frac{p}{N-1}\right)^{N-k}} P(k-1) - \sum_{i=2}^{k} \frac{1 - \binom{N-k+1}{i}\left(\frac{p}{N-1}\right)^{i}\left(1 - \frac{p}{N-1}\right)^{N-k-i+1}}{\left(1 - \frac{p}{N-1}\right)^{N-k}} P(k-i), \qquad k = 2,3,\ldots \tag{6}$$

where the term $P(0)$, i.e. the probability of having an empty destination queue, can be defined to satisfy the following:

$$\sum_{i=0}^{N-1} P(i) = 1 \tag{7}$$

Through some analytical manipulations, the mean switching delay (normalized with respect to τ) spent by a packet in an input queue can be derived as[2,3]

$$T_B = 1 + \sum_{k=0}^{N-1} \frac{(N+k-1)p}{2(N-1)} P(k) + p\,\frac{\sum_{k=0}^{N-1} k[(N-1)(k-1)+p(N-k-1)]}{2(N-1)-p\left[2(N-1)+\sum_{k=0}^{N-1}(N-k-1)pP(k)\right]} + p\,\frac{\sum_{k=0}^{N-1}(N-k-1)(N-k-2)p^2P(k)}{3\left\{2(N-1)^2+p(N-1)\left[2(N-1)+\sum_{k=0}^{N-1}(N-k-1)pP(k)\right]\right\}} \quad (8)$$

Figure 8 shows T_B as a function of p for different values of N achieved under input queueing with the FIFO selection policy. The maximum possible throughput under input queueing for large values of N approaches 0·586.[1,5]

The analytical approach outlined before can be easily extended to consider the case of a finite size for the input buffers. In this case it is evident that depending on p a loss of packets may occur. Figures 9 and 10 show the packet loss probability P_B as a function of p (Figure 9) and of the buffer size (Figure 10). Simulation results have been also reported in these Figures to highlight a good agreement.

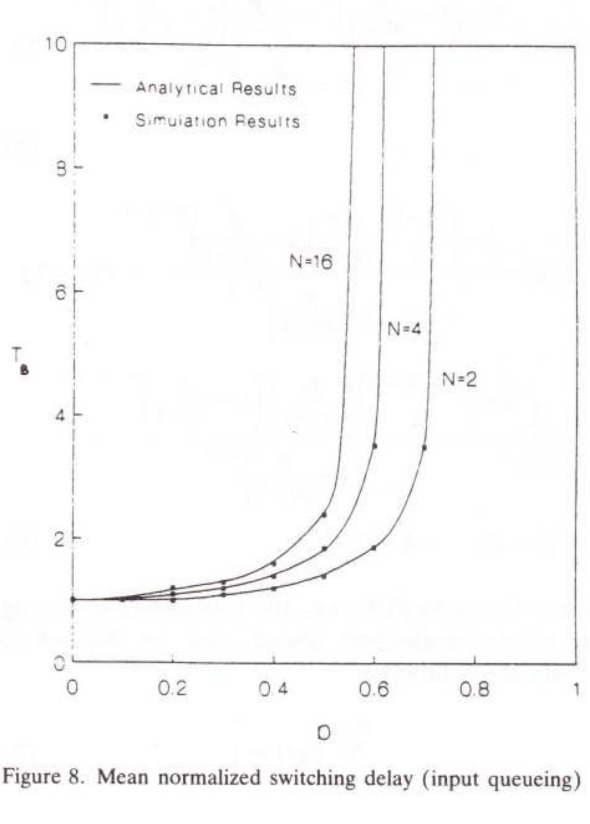

Figure 8. Mean normalized switching delay (input queueing)

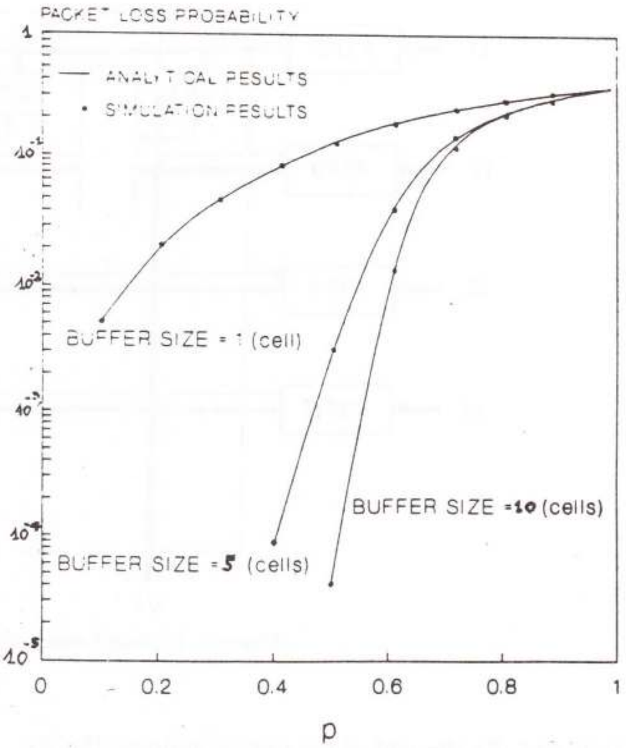

Figure 9. Packet loss probability as a function of p (input queueing)

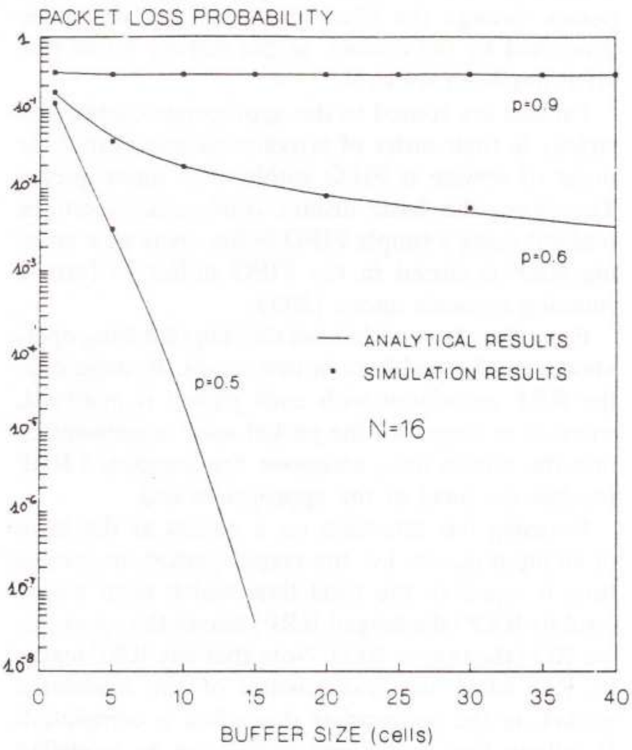

Figure 10. Packet loss probability as a function of the buffer size (input queueing)

Multiple input queueing approach

In the analysis presented above, it has been assumed that none of the packets waiting for routeing in an input queue can leave the queue until the packet at the head has completed service. An efficient alternative to this queueing approach is the multiple queueing on inputs, which always permits routeing for packets in input queues whose RQs are empty.[4,5]

In the multiple queueing approach, each input queue is split into N separate queues, one for each possible down-link (Figure 11). Any new arrived packet is stored in one of these N queues according to its down-link destination.

Packets at the heads of input queues for the same down-link contend for routeing. Assuming again that the service discipline within each input queue is FIFO, it is shown in References 4 and 5 that the analytical approach outlined in the previous section can be applied here again to derive T_B. The final result is

$$T_B = 1 + \sum_{k=0}^{N-1} \frac{(N+k-1)p}{2N(N-1)} P(k)$$
$$+p\,\frac{\sum_{k=0}^{N-1} kN[N(N-1)(k-1)+p(N-k-1)]}{2(N-1)N^2-p\left[2N(N-1)+\sum_{k=0}^{N-1}(N-k-1)pP(k)\right]}$$
$$+p\,\frac{\sum_{k=0}^{N-1}(N-k-1)(N-k-2)p^2P(k)}{3\left\{2(N-1)N^2+p(N-1)\left[2N(N-1)+\sum_{k=0}^{N-1}(N-k-1)pP(k)\right]\right\}} \quad (9)$$

Figure 12 shows T_B as a function of p for different values of N. This Figure highlights that the maximum attainable throughput now approaches 1 as p approaches 1. In Figure 13 the mean switching delay obtained by using the multiple input queueing approach is compared with that obtained by the

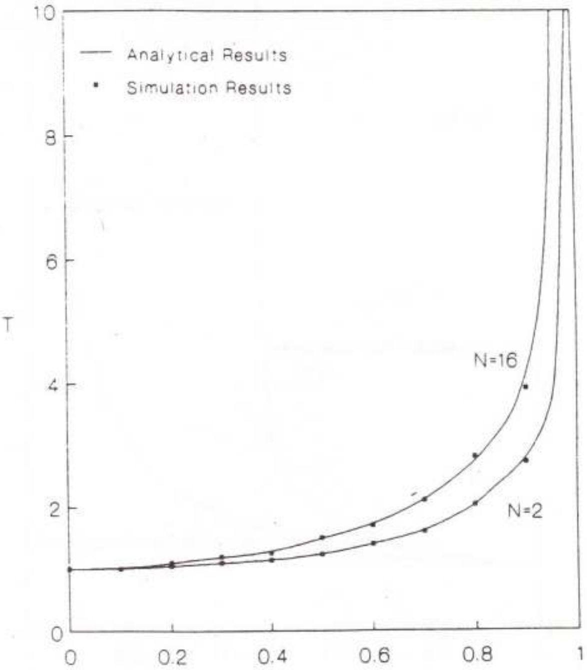

Figure 12. Mean normalized switching delay (multiple input queueing)

single input queueing approach and the output queueing approach which permits optimum throughput mean switching delay performance to be achieved at the expense of an increased implementation complexity.[3] It is evident in this Figure that the proposed multiple queueing on inputs achieves the same performance as the output queueing approach without increasing implementation complexity, i.e. without resorting to a faster switching fabric or having a higher degree of internal connectivity.[3,5]

The analysis presented before can be extended in a straightforward manner to consider the case of

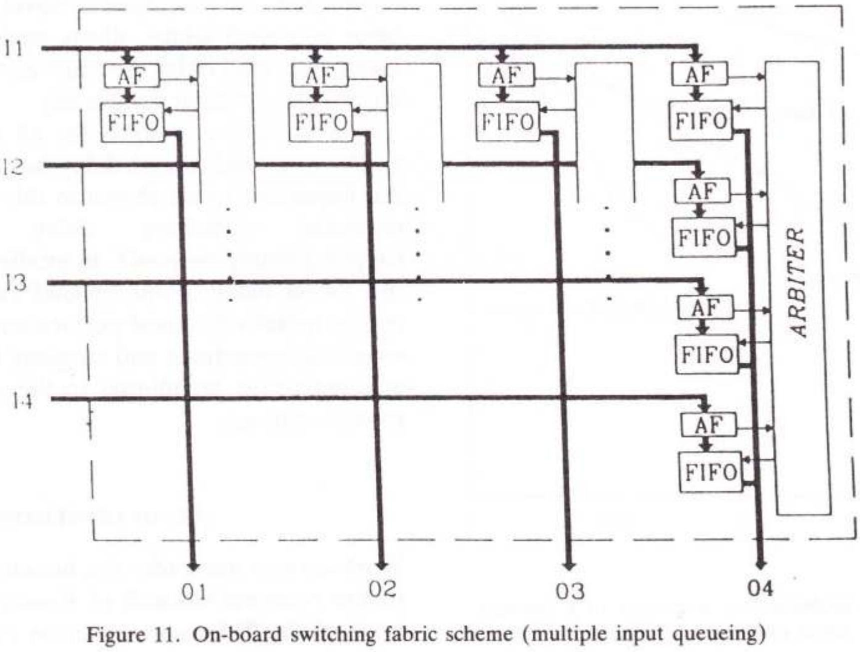

Figure 11. On-board switching fabric scheme (multiple input queueing)

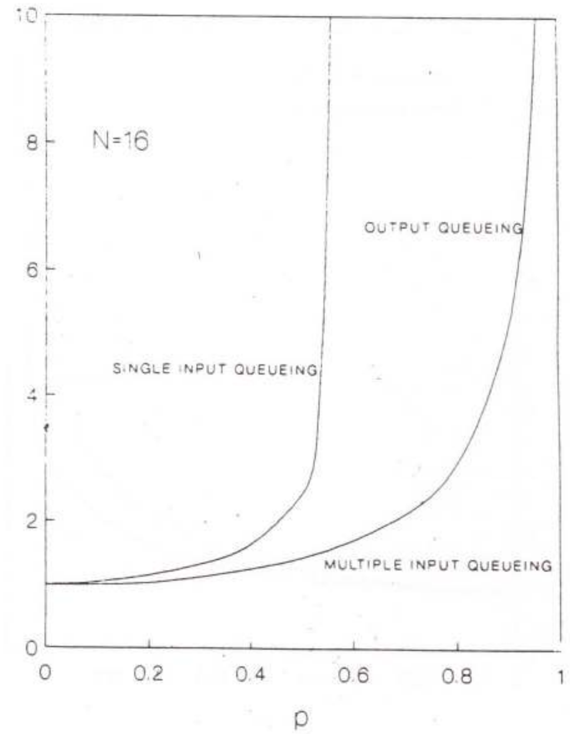

Figure 13. Mean normalized switching delay comparison

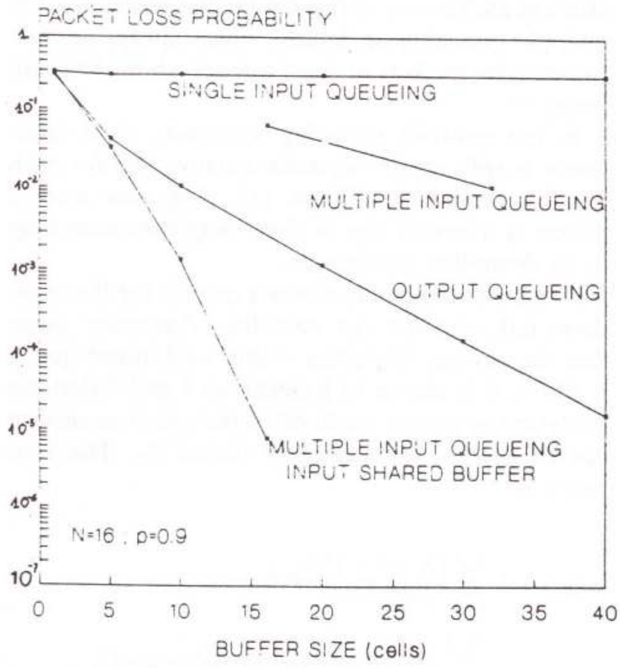

Figure 15. Packet loss probability comparison

buffers of finite size.[5] Figure 14 shows P_B as a function of p for N set to 16 and different values of the buffer size. Comparison with the simulation results is also shown in the Figure. A good agreement is evident.

Better performance can be achieved by considering an implementation architecture of the proposed switching fabric in which all the queued packets at each input share the same buffer.[5] Figure 15 shows P_B in the case of input shared buffers in comparison with that attained by using different queueing approaches. The total buffer size has been assumed as parameter. This Figure highlights that the multiple queueing approach with input shared buffers overcomes all the other approaches.

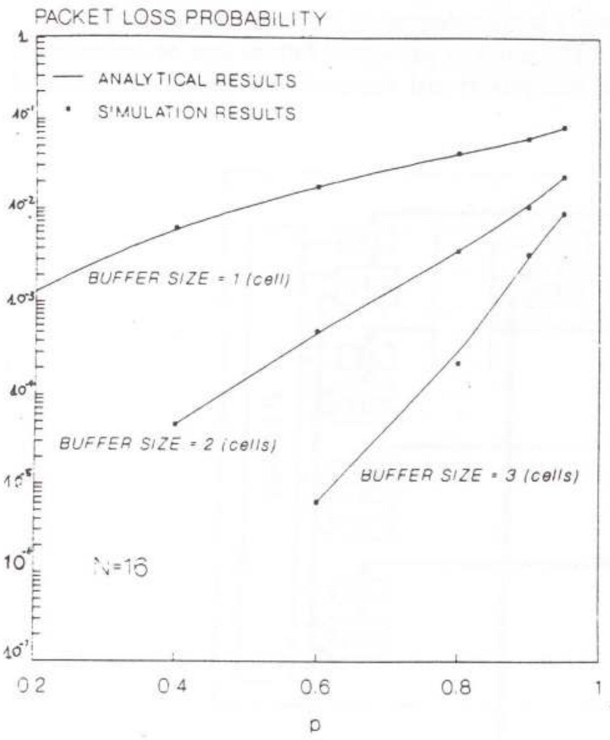

Figure 14. Packet loss probability as a function of p (multiple input queueing)

CONCLUSIONS

In this paper a satellite communication system with on-board fast packet switching capabilities has been described and analysed. Different architectures for the on-board fast packet switching fabric have been considered. In particular a novel architecture has been proposed which allows optimum throughput mean switching delay performance without increasing implementation complexity.

A quantitative analysis for all the contributions to the total end-to-end delay has been performed. An important result shown in this paper is that the on-board switching delay per message (approximately derived), is smaller with respect to the access delay at the ground station (T_U). This results indicated a good performance of the satellite switching operations and suggests looking for more efficient access techniques to the satellite from the ground stations.

ACKNOWLEDGEMENT

Work carried out under the financial support of the Italian National Council of Research in the framework of the Telecommunication Project.

REFERENCES

1. S. Tasaka, 'Multiple-access protocols for satellite packet communication networks: A performance comparison', *IEEE Proceedings*, **72**, 1573–1582 (1984).
2. G. Benelli, E. Del Re, R. Fantacci and F. Mandelli, 'Performance of uplink random-access and downlink TDMA techniques for packet satellite networks', *IEEE Proceedings*, **72**, 1583 (1984).
3. H. Ahmadi and W. E. Denzel, 'A survey of modern high-performance switching techniques', *IEEE J. Select. Areas Commun.*, **7**, 1091–1103 (1989).
4. E. Del Re and R. Fantacci, 'A fast packet switching satellite communication network', *IEEE INFOCOM'91*, Miami, Florida, U.S.A., 7–11 April 1991.
5. E. Del Re and R. Fantacci, 'Performance evaluation of input queueing techniques in ATM switching systems', *IEEE Trans. Commun.*, **COM-41**, (10), 1565–1575 (1993).
6. L. Kleinrock, *Queueing Systems, Vol. 1*, Wiley, New York, 1975.
7. J. F. Hayes, *Modeling Analysis of Computer Communications Networks*. Plenum Press, New York, 1984.
8. J. L. Hammond and P. J. P. O'Reilly, *Performance Analysis of Local Computer Network*, Addison-Wesley Publishing Company, Reading, Massachusetts, U.S.A., 1986.