EXPERIMENTAL RESULTS OF A DEMAND-ASSIGNMENT THIN ROUTE TDMA SYSTEM

Nedo Celandroni, Erina Ferro, Francesco Potortì

CNUCE, Institute of National Research Council
Via S. Maria 36 - 56126 Pisa - Italy
Phone: +39-50-593207/593312/593203
Fax: +39-50-589354 - Telex: 500371
e.ferro@cnuce.cnr.it
n.celandroni@cnuce.cnr.it
F.Potorti@cnuce.cnr.it

SUMMARY

This paper presents the experimental results measured on the Italsat satellite when data is transmitted by using the FODA/IBEA satellite access scheme. The measures presented refer to the packet arrival time jitter, which affects the real-time data, and to the end-to-end delay of the non real-time data. The experimental results, obtained with four Italian active stations on Italsat confirm the analytical results. The paper also shows the advantage of working in *pre-assignment* mode, a feature of the capacity assignment algorithm which improves the end-to-end delay of non real-time data during the transient provoked by a traffic step when the system is scarcely loaded.

KEY WORDS thin route TDMA, real-time data, capacity assignment on demand, jitter, leaky bucket, non-real-time data, end-to-end delay, pre-assignment mode, transient traffic, stationary traffic

1. INTRODUCTION

An efficient and economical system for interconnecting LANs via satellite requires the optimisation of the channel capacity allocation to handle the traffic generated by the users. Aggregated traffic is the characteristic of LANs interconnection traffic. It incorporates real-time traffic (telephony and video) and non real-time traffic (computer data exchange). The real-time traffic is often referred to as *stream,* and includes data from both fixed and variable rate applications. The non-real-time traffic consists typically of *bursty* traffic, classified into two categories: bulk (file transfer) and interactive (TELNET, RLOGIN). Clausen, [1], showed that one of the interesting results in investigating LAN traffic, particularly in a heavy loaded condition, is that delays are at least as long as the satellite propagation delay, mostly due to queueing delays in different routers and bridges.

Preprint of an article published on the INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS, VOL. 14, No. 2

Therefore, it is important to measure the queueing delay which is added to the data by the satellite access scheme used (in addition to the unavoidable round trip time), when the satellite link is included in the interconnection. Moreover, bad weather conditions during the data transmission via satellite have a heavy impact in terms of quality of service and delivery delay. The literature presents many adaptive methods to counteract rain attenuation at frequencies above 10 GHz in satellite communication systems [2, 3], but none of them can simultaneously satisfy the following requirements:

- simultaneous transmission of stream and bursty data,
- maintenance of the quality of service rigorously close to user requirements,
- channel optimisation, in any weather conditions,
- cost efficiency,
- robustness.

Fade countermeasure solutions such as Frequency Diversity or Site Diversity are obviously very expensive due to the duplication of the hardware used. Solutions in TDMA are certainly the most flexible way of allowing multiple earth-stations to share the capacity of a given channel. Only a few of them have actually been implemented as prototypes and even less have entered production. The American ACTS system [4, 5] was developed to become the standard of communication satellites of the future. It provides demand assignment multiple access digital communications, two types of on-board switching, wide bandwidth channels, rain fade compensation and Ka band propagation beacons. Over 45 different experiments have been carried out, some of them are still in progress. The ACTS program has paved the way for the use of hopping spot beams and demand assigned on-board switching, but it is impossible to compare its performance with the FODA/IBEA system, which considers the satellite only as a passive bent pipe channel. It is generally very difficult to compare the performance of different systems, mainly as a result of the lack of a common testbed for the various solutions.

FODA/IBEA(1) was designed to satisfy the requirements listed above, and the whole system (TDMA controller hardware plus the access scheme software) is a prototype used in the Italian experiment of the LAN interconnection via satellite (*Thin Route TDMA for LANs Interconnection* project). The fade countermeasure technique adopted adapts the energy per information bit to each individual link status, which depends on atmospheric conditions. The total attenuation of each link (up-link plus down-link) is compensated for by varying the transmitting power, the data coding and bit rates. The fade countermeasure technique adopted exploits a prototype of a modem, developed by GEC Marconi R. C. (U.K.) which can dynamically change the data bit rate. The price of this feature is the need for a short preamble between sub-bursts, besides the rather long preamble needed before each burst to allow carrier and bit timing synchronisation [6].

The aim of this paper is to present experimental results of the satellite access scheme performance measured on the Italsat satellite. We focused on measuring the most significant parameters, i.e. the

(1) Fifo Ordered Demand Assignment/Information Bit Energy Adaptive

delay and the delay variation (jitter), introduced by the satellite link on the data transmitted. The paper is organised as follows. Section 2 describes the capacity allocation policy for the stream and the bursty data. There is a brief reminder of the jitter theory, together with some results on the analytical study carried out on the bursty data end-to-end delay. Section 3 describes the environment where the metrics selected were measured. The jitter was measured in various traffic loading conditions, showing the agreement with theory. The measures relevant to the end-to-end delay of the bursty data, both in the stationary and in the transient cases are also presented. Conclusions are drawn in Section 4. In references 15÷23 other access schemes are reported for comparison.

2. THE CAPACITY ALLOCATION POLICY

In FODA/IBEA a master station is responsible for system synchronisation and for capacity allocation on demand of the traffic stations. These tasks are accomplished by sending a reference burst which contains the burst time plan (BTP), at the beginning of each 20 ms *frame*. The BTP is the transmission window layout allocated on the traffic stations' request basis. A maximum of one window is allocated for each requesting station, in order to save the overhead due to the rather long preambles needed by the modem for the burst synchronisation. Inside the transmission window the traffic stations' multiplex data comes from both stream and bursty applications. Data packets generated by each application are sent in sub-bursts, by adapting the coding and bit rates to the current condition of each individual link and to the BER required by each application. Broadcast and multicast transmissions are also supported. In this case the worst link status is considered to adapt the transmission parameters.

In faded conditions, the channel capacity reserved for bursty transmissions is reduced up to a minimum value (which may be set to zero), in order to maintain the stream sessions already set up.

The transmission bit rate can vary from 1 to 8 Mb/s, and the coding rates used are: 4/5, 2/3, 1/2 and uncoded.

2.1 Stream Traffic

Stream capacity is guaranteed by the system, i.e. once the request has been accepted the assignment is maintained until released by the user. The total stream capacity is allocated up to a maximum value; the remaining amount of the capacity is devoted to the bursty data. The algorithm for the bursty capacity allocation does not depend on the presence of stream data in the frame.

Any stream capacity request granted involves the allocation of one transmission window per frame. In the same transmission window, after the stream allocation, space is allocated for any bursty transmission, if the station made a request for it. This part is not regular in the frames due to the bursty assignment algorithm adopted. This means that the transmission window assigned to a station may vary either in size or in the position inside the frame, due to the variability of the assignments of the previous stations.

As a consequence of the allocation policies and fade handling, when a sequence of packets crosses the satellite network, each packet may experience a different delay. This delay variation induces a jitter [7] on the packet arrival times. While this jitter is almost irrelevant for non-real time applications, it may cause problems for the real-time traffic because the regularity of the packet generation time intervals is no longer respected. In order to absorb the jitter, the packets which arrive with a shorter delay must therefore be put into a buffer (*leaky bucket*) to wait for packets with a longer delay. The packets are parked in this buffer for a convenient amount of time, in order to guarantee either a continuous flux of output data or to avoid an overflow, when delayed or early packets arrive, respectively. The penalty for jitter absorption is the introduction of an additional end-to-end delay, i.e. the time spent by the data inside the buffer.

2.1.1 The jitter theory

Below we recall the jitter theory to show the agreement of the experimental results with the theory. Let us consider a simple system where a transmitting stream application converts a continuous, constant rate stream of bits into a sequence of packets equal in length and equally spaced. The stream assignment for each application consists of a number of bytes per frame which is generally different from the packet length. Therefore, when packets cross the satellite network, they may be disassembled or multiplexed at the system input and reassembled or demultiplexed at the system output, generally losing their time relationship. The receiving application needs a leaky bucket for the incoming packets to reconstruct the original constant rate bit stream. After transmission to the satellite, the packets generated by the transmitting application arrive at the leaky bucket, which reproduces a plesiochronous stream of bits with a mean rate equal to the sender rate. We assume that the four communication lines that connect the boxes in Fig. 1 have infinite speed. Another assumption, which will be removed later on, is that the transmission window assigned by FODA/IBEA is in a fixed position inside the frame. The output of FODA/IBEA is thus a sequence of packet bursts which are spaced out by an integer number of frames.

Fig 1. Block view of the system under study.

Let us now define the jitter of the packets as the difference between the time they leave FODA/IBEA and the time they would be expected to arrive if their time relationship were not changed by the satellite network crossing. In other words, if packets enter FODA/IBEA with an inter-arrival time equal to δ, the first packet leaving FODA/IBEA at time t_1 has, by definition, a null jitter ($j_1 = 0$), while the n^{th} packet (coming at the time t_n) has the jitter j_n given by:

$$j_n = t_n - t_1 - (n-1)\delta . \qquad (1)$$

The packets leaving FODA/IBEA have a comb distribution of the jitter made up of N lines of equal height $1/N$, equally spaced by a T_f / N time, where T_f is the frame length. If the ratio δ / T_f

can be expressed as a rational number m/n where m and n are natural numbers, then $N = n/GCD(m, n)$. The distance between the lines farthest apart, i.e. the comb span J, is:

$$J = j_{\max} - j_{\min} = T_f(1 - 1/N) \qquad (2)$$

If the ratio δ/T_f cannot be expressed as a rational number, then the jitter distribution has a constant density in a range whose span is T_f.

The comb can be displaced anywhere on the jitter axis, as long as one of its teeth lies on the origin of the axis. In fact, the jitter of all the packets after the first one depends on the arrival of the first packet. Since the jitter sequence is periodic, any packet in the sequence can be taken as the first one, so nothing can be said about the actual distribution displacement for an arbitrary sequence of packets [8].

Let us call *delay offset* the time introduced by the leaky bucket between the first packet's arrival and the beginning of the output data flow [9]. If a packet arrives at the buffer input while there is not enough free space in the buffer, it is discarded and its data lost (*buffer overflow*). In case the buffer empties, its output stops until the next packet is received at the input (*buffer underflow*).

Knowledge of the jitter distribution span at the leaky bucket input, J, is sufficient to set the buffer size and the delay offset in the worst case. Let us denote by B_{fn} the buffer fullness (expressed in number of packets) at the arrival of the n^{th} packet, and by D_o the delay offset. If the buffer neither overflows nor underflows we have:

$$B_{fn} = n - (t_n - t_1 - D_o)/\delta \qquad (3)$$

Substituting (1) in (3) we get:

$$B_{fn} = (D_o - j_n)/\delta + 1 \qquad (4)$$

To avoid buffer underflows, for any n, we must have $B_{fn} \geq 1$, because data are available in the buffer if at least one packet has been completely received. From the above $D_o \geq j_n$, so D, the minimum value of D_o, is equal to:

$$D = j_{\max} \qquad (5)$$

The minimum buffer size B to avoid any overflow is equal to the maximum buffer fullness. Substituting (5) in (4) we have:

$$B = B_{f\max} = (j_{\max} - j_{\min})/\delta + 1 = J/\delta + 1 \qquad (6)$$

In general, however, $j_{\max}$ and $j_{\min}$ are not known, but only J is known. This leads us to set D and the buffer size as:

$$D = J$$
$$B = 2J/\delta + 1 \quad \text{[packets]} \qquad (7)$$

Let us now compute the worst case estimate of the jitter distribution span in reality. An unavoidable clock difference exists between the application and the FODA/IBEA clocks, which causes the comb distribution to eventually evolve into a uniform distribution with a span $J = T_f$. Furthermore, there are two reasons why the jitter distribution span may become greater than T_f. The first is that if the packets entering FODA/IBEA are already affected by a jitter, whose distribution span is J_i, this quantity must be added to J [8]. The second reason is because the transmission window of any given application is not in a fixed position inside the frame. This causes a further jitter, whose size is less than T_f.

In the worst case (denoted by the subscript w), the jitter span, the buffer size and the delay offset thus become respectively:

$$J_w = 2T_f + J_i$$
$$B_w = 1 + 2(2T_f + J_i) / \delta \qquad \text{[packets]} \qquad (8)$$
$$D_w = 2T_f + J_i$$

2.2 Bursty traffic

While the delay of stream traffic must be kept as low as possible and with constant packet inter-arrival times, the delay of bursty traffic is not so critical. In any case it must be investigated in order either to tune-up efficient application protocols or to implement suitable congestion control mechanisms, which are required to improve system stability.

The request for a bursty transmission window, sent by a traffic station, is proportional to the traffic i coming into the station plus the backlog q, i.e. the volume of data waiting for transmission on satellite. We have:

$$r = q + H\, i \qquad (9)$$

where H is a temporal constant of proportionality. Simulation results, obtained by loading the channel with Poisson generators of bursty traffic for 10 stations (Fig. 2), suggested using 0.4s as the H parameter.

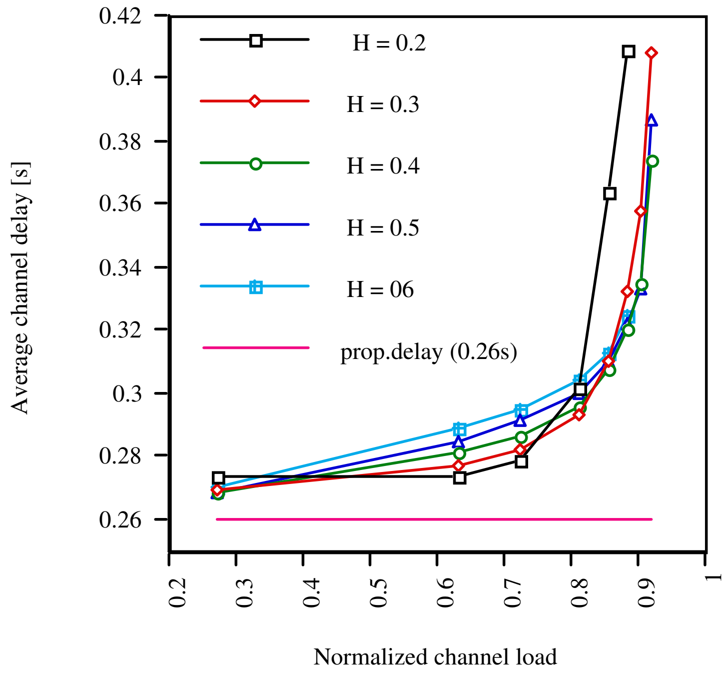

Fig. 2. Average end-to-end delay (10 stations), versus channel load, for various values of H. The system is loaded with Poisson generators. The channel load includes the packet headers.

Bursty data requests are issued as frequently as possible, to update the master station about the latest changes in the traffic situation. The requests are piggy-backed with the data, when an assignment already exists for the requesting station, otherwise they are sent in a control window whose assignment is guaranteed with a minimum frequency.

The requests are organised by the master into a circular queue which is cyclically scanned to compute the amount of time for each assignment. Any bursty request received from the same station replaces the previous value. The length of the assigned transmission window w is proportional to the request in a range of values between a minimum $(w_{\min})$ and a maximum $(w_{\max})$. We have:

$$w = \min(w_{\max},\ \max(w_{\min},\ f\ \ r)). \qquad (10)$$

The coefficient f was chosen, for the current implementation, equal to the number of active stations N divided by 100, with a minimum of 0.05 and a maximum of 0.5. The threshold $w_{\min}$ was introduced for efficiency purposes. It prevents the information part of the allocation from being too small in comparison with the transmission overhead, due to preambles and headers. The threshold $w_{\max}$ prevents a heavy loaded station from removing too much capacity from the other stations.

After each assignment, the bursty request is decreased by the assignment itself, and the next request is analysed, if space is still available in the frame. The first assignment that does not entirely fit the current frame is considered as the first in the next frame, where the rest of the computed amount is assigned. All the space up to the end of the frame (if insufficient for a minimum assignment) is given as an over-assignment to the last station processed.

Any space available in the frame after an entire assignment cycle, i.e. the time between two consecutive allocations to the same station, is shared among all the active stations, including those stations which had no bursty assignment in that frame. The system behaves like a sort of fixed TDMA (F-TDMA), called *pre-assignment mode*, which is accentuated in inverse proportion to how much the system is loaded. When the channel load increases beyond a certain limit, called *pre-assignment limit,* the F-TDMA component is extinguished and the assignment algorithm strictly follows (10).

2.2.1 Notes on the analytical study

In [10] the bursty assignment algorithm was studied under linearity conditions. The study shows that when the system is loaded with fixed rate traffic generators all the stations have a null steady state backlog if :

$$\sum_{j=1}^{N} i_j \leq C_d \tag{11}$$

where i_j is the traffic entering the j^{th} station, N is the number of active stations and C_d is the channel capacity left to bursty data.

Still considering stationary conditions of the system, let us denote by the assignment cycle and by T_f the frame length. The system works in pre-assignment mode when some residual capacity C_r is still available in the frame after a complete assignment cycle. In this case $T_a = T_f$ and each active station receives an extra assignment p equal to C_r / N. The system works in pre-assignment mode until:

$$C_r = C_d - \frac{f}{T_f} \sum_{j=1}^{N} r_j > 0 \tag{12}$$

Substituting (9) in (12), under the hypothesis of a null backlog at each station, the pre-assignment condition becomes:

$$\sum_{j=1}^{N} i_j < \frac{T_f}{f\ H} C_d \, . \tag{13}$$

With the chosen minimum value of f equal to 0.05 and the H factor equal to 0.4 s, we have $\frac{T_f}{f\ H} = 1$. Therefore, the pre-assignment limit is reached when the total load equals C_d.

In pre-assignment mode each station has a $1/N$ of the spare capacity pre-assigned to absorb immediately modest transients of traffic, while up to the entire spare capacity is devoted to absorbing bigger transients of traffic with the allocation mechanism. However, as it is reasonable to think that each station traffic has an order of magnitude proportional to $1/N$, when the system works with a high number of stations the condition $\frac{f \; H}{T_f} = 1$ may cause too many allocations to fall below $w_{\min}$ with a consequent compression of the system dynamics and waste of capacity. This is why we decided to make $f \; H$ proportional to N. Specifically, once the H factor was fixed as the best result from the simulation, we chose $f = \max\,(0.05,\; \min\,(N/100,\; 0.5))$. With $H = 0.4$ and $T_f = 0.02$ we always have $\frac{f \; H}{T_f} \geq 1$.

The original FODA system [11], which did not support the fade countermeasure feature, did not use the pre-assignment mode. In fact, any spare space after the assignment cycle was only given to the stations which already had an allocation. In [10] the convenience in working in pre-assignment mode is shown, during a transient of traffic which affects a station, when all the other stations are in a steady state condition with a null backlog. The transient case was chosen because it is the most critical one from the end-to-end delay point of view. The conditions for the validity of the linear analysis are:

- the amplitude of the traffic step at the station considered is small enough to make the relative increment of T_a negligible

  or

- in pre-assignment mode, the assignment to the station considered is not bigger than C_r for the entire duration of the transient.

Under the above conditions the assignment can be considered as being proportional to the request during the whole transient. As a result, only the delay and the queue length relevant to the station which experiences the traffic step varies with the time, while all the other stations are not affected by that event.

The reliability of the model used is shown in Fig. 3 where the analytical results are compared with the measured ones. Before the transient step of 1 Mbit/s, the station considered is loaded with 1 Mbit/s traffic. The overall channel load is such that the second of the above linearity conditions is fully respected.

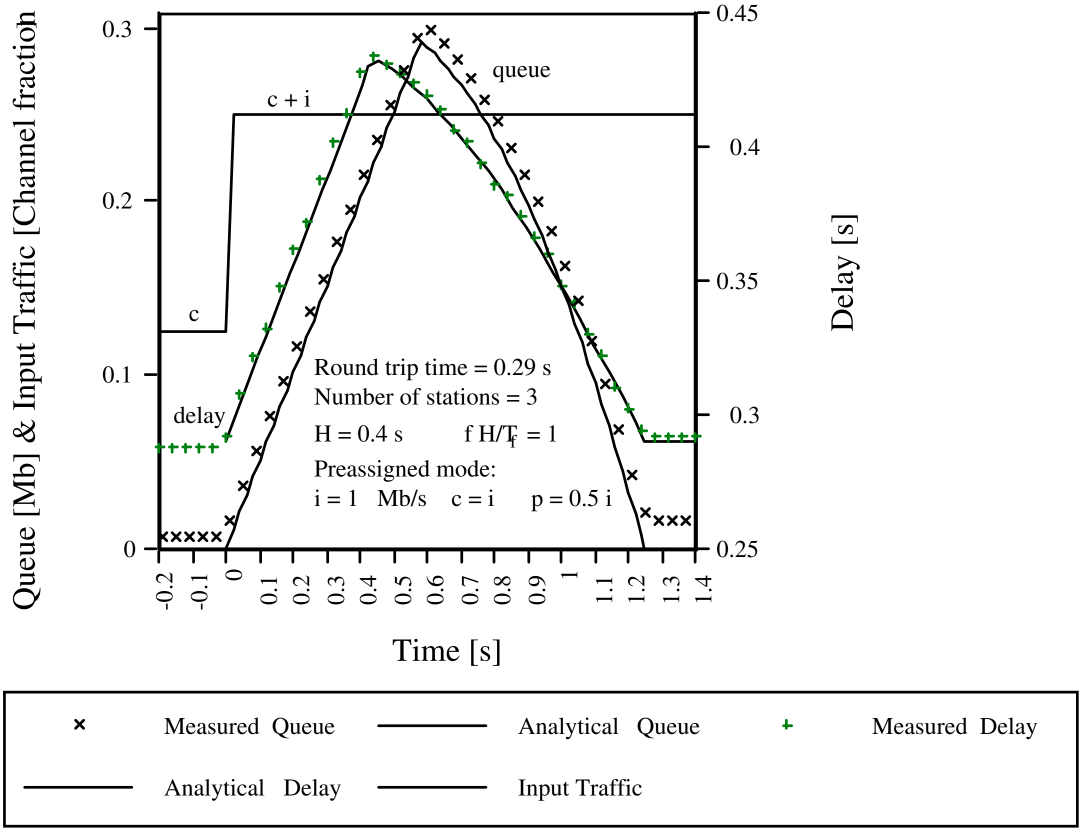

Fig. 3. Queue and delay versus time during a transient due to a 1Mbit/s traffic step at one of the three working stations. Comparison between analytical and experimental results.

Figure 4 shows the convenience of working in pre-assignment mode. The figure compares the end-to-end delay as a function of time during a transient, when the system is and is not working in pre-assignment mode, respectively. Two conditions are analysed: a) before the traffic step, the station has a constant load of amplitude c equal to the amplitude of the traffic step i; b) before the traffic step the station has no previous load (case $c = 0$). The gain of the non pre-assignment mode in case $c = i$ is clearly modest (the maximum delay is the same in both cases) in comparison with the gain in pre-assignment mode when $c = 0$. In [10] we demonstrated that this fact is more evident with lower loading conditions in the system.

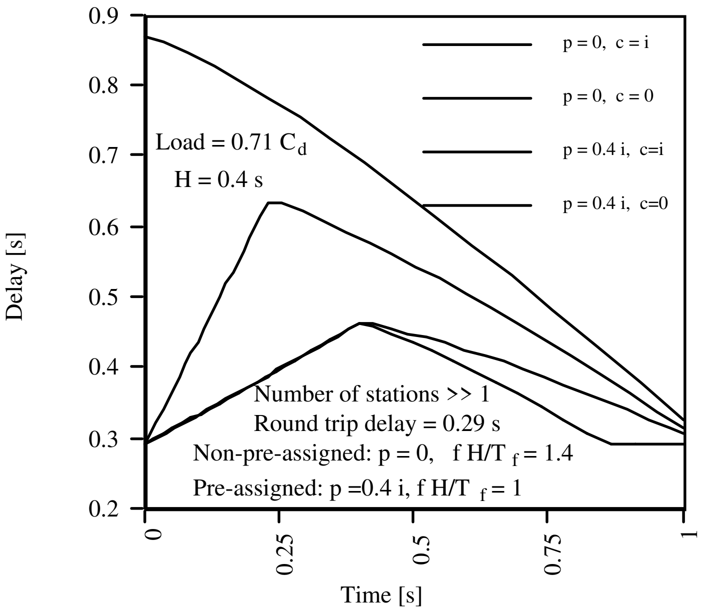

Fig. 4. End-to-end delay versus time during a transient due to a step of traffic at one station. A large number of stations is considered. Pre-assignment and non pre-assignment cases for pre-loaded and for non pre-loaded stations are reported.

3. EXPERIMENTAL RESULTS

In 1994 FODA/IBEA was used to interconnect at 8 Mbit/s via the Italsat satellite three earth stations in the framework of the "Thin Route TDMA for LANs Interconnection" project, carried out in collaboration with the University of Florence [12]. Two stations were located in Pisa, while the third one was located in Florence (about 90 Km away). Although for practical reasons at that time the experiment only involved three stations, this configuration could easily be extended to a larger number of stations, up to the hardware limit of 112. Each station needed a gateway that connected it to an Ethernet LAN and to stream applications (data, voice, video). In general, the stations can themselves be Ethernet hosts, provided that the applications that use the satellite link are equipped with the GAFO(2) protocol [13], an ad-hoc developed interface for handling the data exchange with the FODA/IBEA system. In the configuration used in the experiment, stream applications run on hosts connected to a Token Ring, while bursty applications run on hosts connected via Ethernet. Data from both the LANs converged at the gateway which communicated with the FODA/IBEA system by using the GAFO protocol. The experiment was successful. It demonstrated the high quality and reliability of the satellite network developed, but the measures collected did not present any actual interest because the traffic loads of the LANs involved in the interconnection were not able to stress the system. Therefore, we made measures on Italsat by using four stations all fed by a traffic generator, named MTG(3) [14], instead of by real traffic. MTG can generate data packets on

(2) GAteway-FOda/ibea
(3) Multiapplication Traffic Generator

Ethernet according to a chosen distribution and record them onto disk, after being looped-back by the satellite network. The use of MTG allowed us to stress the system as much as we wished in order to highlight particular aspects of its behaviour. When the measures were made, Poisson traffic was still considered as an acceptable model for LAN traffic, therefore, at that time we did not use any fractal traffic generator, which is now considered the best model for the LAN traffic. At the current time, although we have implemented such a traffic generator on MTG, we cannot make any additional measures on satellite.

3.1 Experimental results on the jitter

The time jitter was calculated by measuring the difference between the MTG receiving times and the expected times of the data packets.

The value of the system time frame was 20 ms. Applications generating packets with an inter-arrival time less than, equal to, or greater than the frame length, respectively, were considered. The traffic pattern of three classes of applications is shown in Tabs. 1, 2 and 3. Three stream and three bursty traffic generators were considered in all the experiments, but some measures were made without sending bursty traffic to the channel. The arrival at the FODA/IBEA system of data already affected by a jitter due to the various LANs crossing was simulated by introducing an initial jitter with a uniform distribution. The span value of this initial jitter is indicated in the tables.

All the tests were performed at 8 Mbit/s, with uncoded data and in clear sky conditions. All the measures are relevant to the data generated by the traffic generator No. 3 (TG3). In each test, more than 60,000 packets were collected.

TG number (application)	Data type	Throughput [Kbps]	Packet length [bytes]	δ [ms]	Source	Destination	Data	Maxi initial jitter [ms]
1	stream	341	128	3	1	2	fixed	2
2	stream	171	128	6	2	2	fixed	5
3	stream	29	128	35	3	3	fixed	34 24 0
4	bulk	800	300	3	1	2	poisson	0
5	bulk	400	300	6	2	2	poisson	0
6	bulk	120	300	20	3	2	poisson	0

Tab. 1. Traffic environment for the jitter measurement of application No. 3. Data packets with $\delta = 35$ ms.

TG number (application)	Data type	Throughput [Kbps]	Packet length [bytes]	δ [ms]	Source	Destination	Data	Max initial jitter [ms]
1	stream	341	128	3	1	2	fixed	2
2	stream	171	128	6	2	2	fixed	5
3	stream	51	128	20	3	3	fixed	200
4	bulk	800	300	3	1	2	poisson	0
5	bulk	400	300	6	2	2	poisson	0
6	bulk	120	300	20	3	2	poisson	0

Tab. 2. Traffic environment for the jitter measurement of application No. 3. Data packets with $\delta = 20$ ms.

TG number (application)	Data type	Throughput [Kbps]	Packet length [bytes]	δ [ms]	Source	Destination	Data	Max initial jitter [ms]
1	stream	341	128	3	1	2	fixed	2
2	stream	29	128	35	2	2	fixed	34
3	stream	171	128	6	3	3	fixed	60
4	bulk	800	300	3	1	2	poisson	0
5	bulk	400	300	6	2	2	poisson	0
6	bulk	120	300	20	3	2	poisson	0

Tab. 3. Traffic environment for the jitter measurement of application No. 3. Data packets with $\delta = 6$ ms.

Figures 5÷14 show the jitter distributions measured in a variety of tests. All the distributions lost the theoretical comb shape, even in the tests in which no bursty data and no initial jitter were present. As stated above, this is due to the difference between the application clock and the FODA/IBEA clock. This difference (about 1.5 10^{-5}) causes the comb distribution to drift. If the total drift offset accumulated in the observation time is greater than the distance between two close teeth in the comb, the distribution appears as continuous in an interval equal to T_f.

Figure 11 highlights that in 20,000 samples time the teeth of the comb are still separate, while in 60,000 samples time they are not.

The buffer size and the delay offset to avoid any overflow and underflow can be deduced, for each run, by taking the values j_{min} and j_{max} on the abscissas of the distribution diagrams and applying relations (5) and (6).

Setting the size of D and B according to relations (5) and (6) only gives valid results for the run considered. Different runs, performed with the same traffic conditions, produce the same jitter distribution shape, but its displacement may be shifted on the abscissa, depending when the first packet arrives. Relation (7) must then be used for the actual sizing of the D and B parameters. Relation (7), applied on measured results of the jitter span, does not ensure complete safety in avoiding buffer overflows or underflows, because the measurements are made with a finite number of samples. However, if this number is sufficiently large, as in the cases presented, the use of relation (7) instead of the safe but fairly conservative relation (8), may be very acceptable in practice, thus reducing the sizes of D and B. The reduction in the delay offset is particularly advantageous for applications such as phones or video conferences, because it reduces the packet end-to-end delay.

In the most general cases bursty data are present on the channel as well, and stream data are affected by an initial jitter. For these cases measurements were made to show how many buffer underflows were caused by undersizing the delay offset, in the hypothesis that buffer overflows must be avoided (i.e. the buffer size is kept equal to the optimal size). Figures 15, 16 and 17 show the results of these measurements for the 35 ms, the 20 ms and the 6 ms cases, respectively.
It can be observed [9] that if $D < j_{max}$, the system may receive late packets until it receives the packet with the maximum jitter. Afterwards, the situation is equivalent to the one in which a delay offset $D = j_{max}$ has been introduced at the beginning and no more packets arrive late.

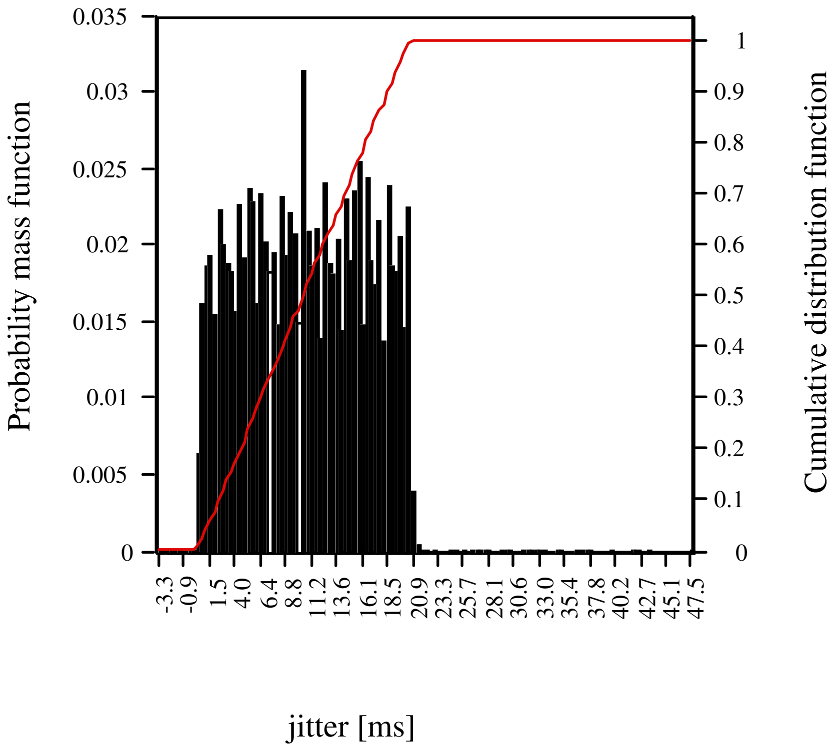

Fig. 5. Jitter distribution for δ = 35 ms; bursty data present on the channel; no initial jitter.

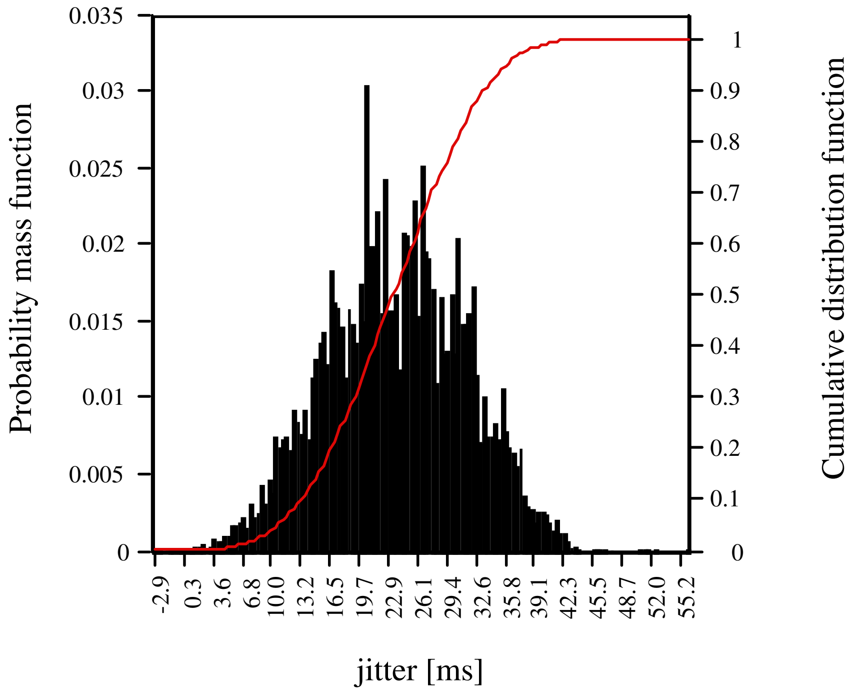

Fig. 6. Jitter distribution for δ = 35 ms; bursty data present on the channel; initial jitter span = 34 ms.

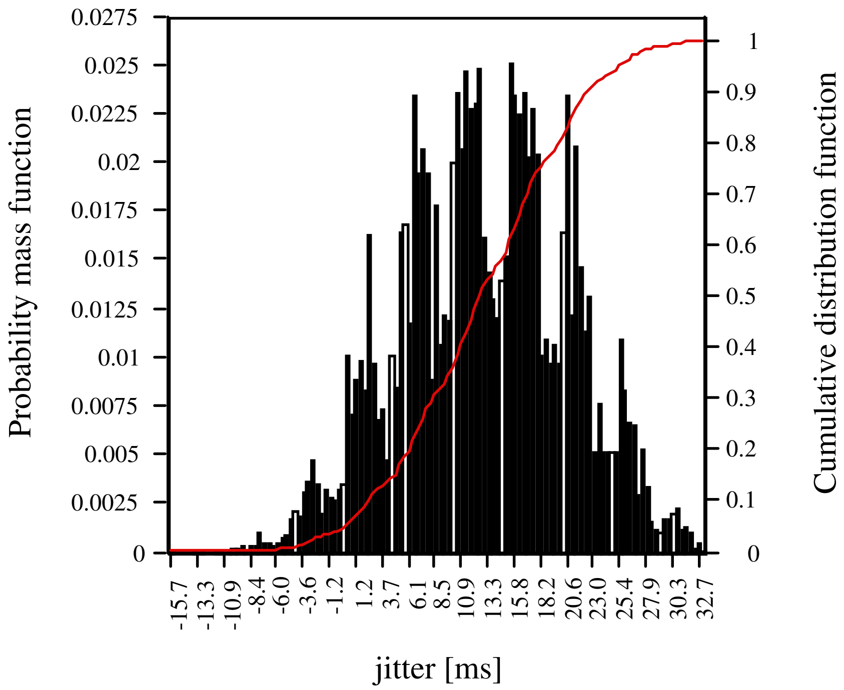

Fig. 7. Jitter distribution for δ = 35 ms; no bursty data present on the channel; initial jitter span = 34 ms.

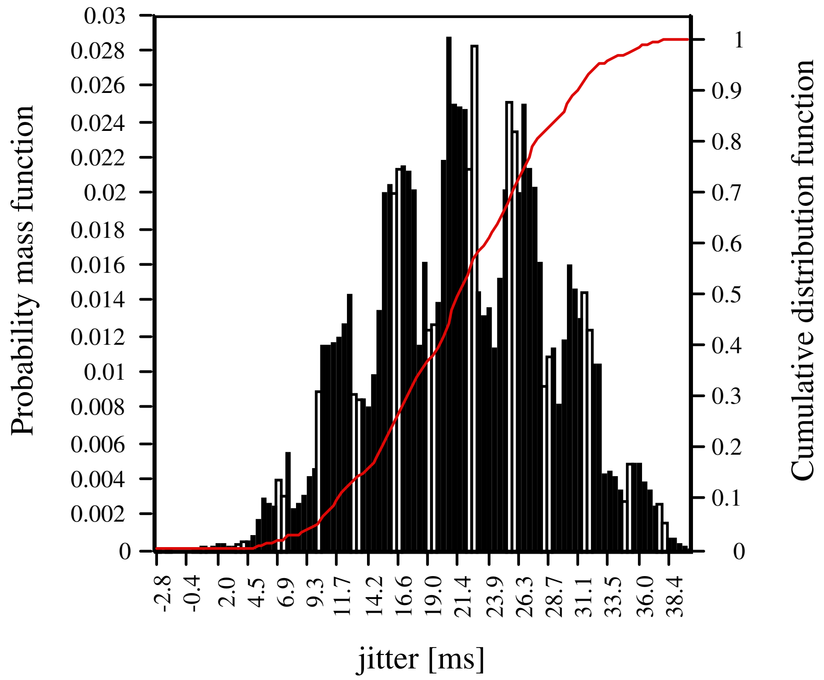

Fig. 8. Jitter distribution for δ = 35 ms; no bursty data present on the channel; initial jitter span = 24 ms.

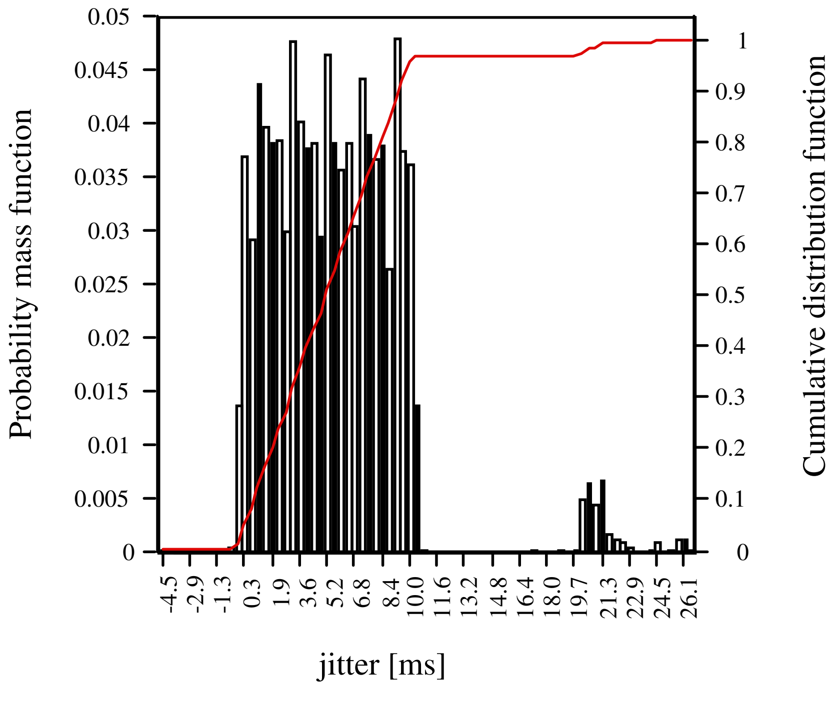

Fig. 9. Jitter distribution for δ = 20 ms; bursty data present on the channel; no initial jitter.

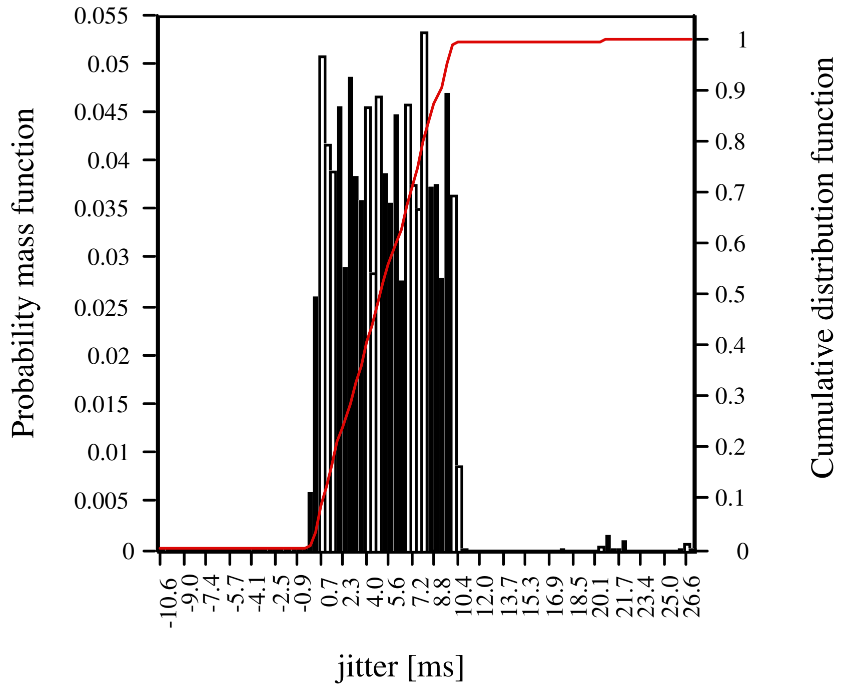

Fig. 10. Jitter distribution for $\delta = 20$ ms; bursty data present on the channel; initial jitter span = 19 ms.

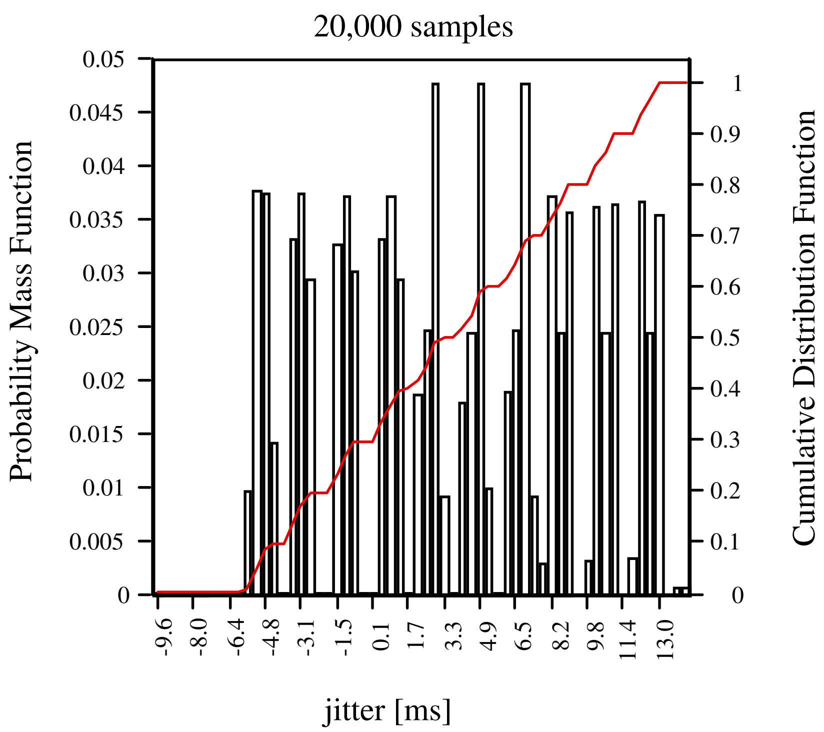

Fig. 11a. Jitter distribution for $\delta = 6$ ms; no bursty data on the channel; no initial jitter. 20,000 samples.

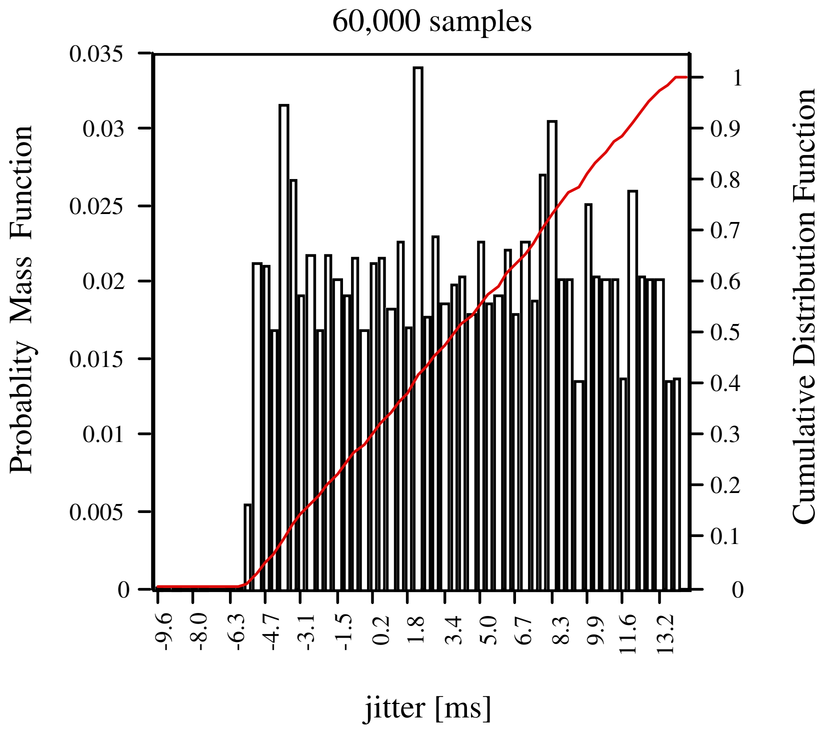

Fig. 11b. Jitter distribution for $\delta = 6$ ms; no bursty data on the channel; no initial jitter. 60,000 samples.

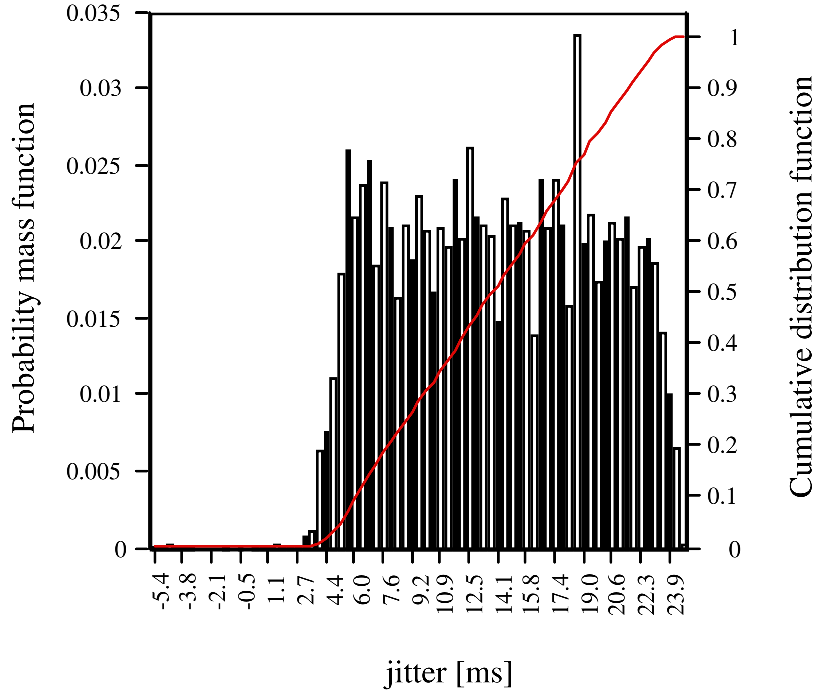

Fig. 12. Jitter distribution for $\delta = 6$ ms; no bursty data present on the channel; initial jitter span = 6 ms.

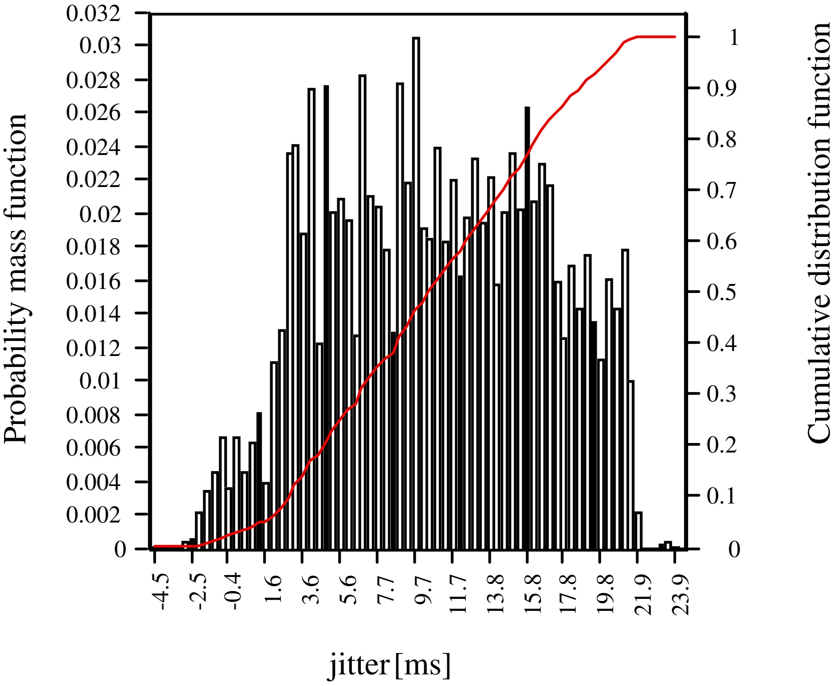

Fig. 13. Jitter distribution for δ = 6 ms; bursty data present on the channel; no initial jitter.

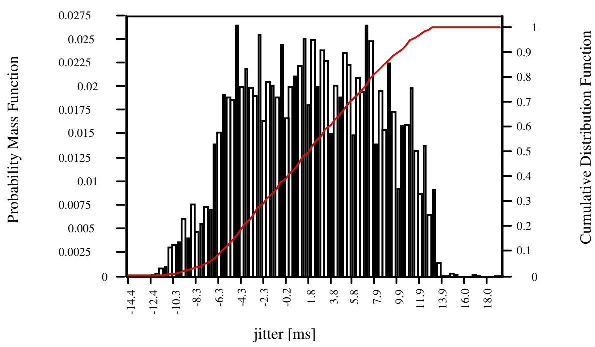

Fig. 14. Jitter distribution for δ = 6 ms; bursty data present on the channel; initial jitter span = 6 ms.

Fig. 15. Late packets versus delay offset; δ = 35 ms; bursty data present on the channel; initial jitter span = 34 ms.

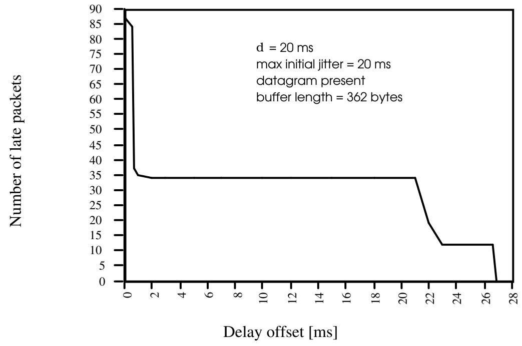

Fig. 16. Late packets versus delay offset; δ = 20 ms; bursty data present on the channel; initial jitter span = 20 ms.

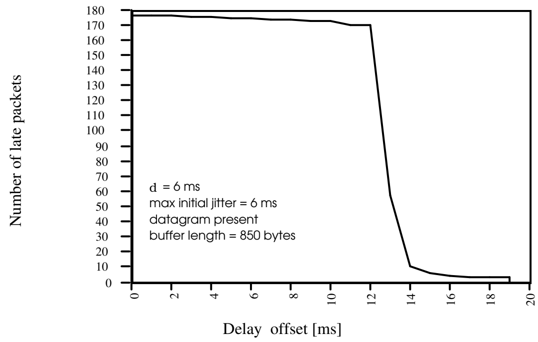

Fig. 17. Late packets versus delay offset; δ = 6 ms; bursty data present on the channel; initial jitter span = 6 ms.

3.2 Experimental results on the bursty data end-to-end delay

Note that for implementation reasons data was sent at an average of one frame later from its arrival at the TDMA controller. Including some processing overheads, the satellite round trip time must be considered as being around 280 ms, while the net satellite round trip delay for Italsat at the Pisa station is 253 ms. The end-to-end delay shown in all the diagrams includes the satellite round-trip-time and the queuing time that data must wait for before being transmitted.

In all the measures data was sent uncoded at 8 Mbit/s, the reference burst was sent 1/2 coded at 2 Mbit/s, and any control slots were assigned at 1 Mbit/s. No stream capacity was allocated. Moreover, in all the figures the traffic loads indicated are information data loads, and no overheads due to preambles, headers, signalling bursts and guard times are included. They are expressed as a fraction of the total channel capacity (8,192 Kbit/s).

The length of the packets is between 512 and 1024 bytes for the stationary cases and 1024 bytes for all the transient cases.

3.2.1 The stationary case

Figures 18 and 19 show the transmission end-to-end delay experienced by each of the four active stations as a function of the traffic handled by the station. The five dashed lines show different runs with different channel loading conditions, while the continuous lines show the behaviour of each single station.

The traffic transmitted was not distributed evenly among the stations to allow comparisons of behaviour among different loaded stations in various channel loading conditions.

The diagrams show the steady state behaviour of the system when the traffic sent by each station is fixed rate or Poisson, respectively. The delay values are averaged over a period of 30 s, starting 15 s after the beginning of the transmission to exclude the transient effects.

The same runs that produced Figs. 18 and 19 were used to show the overall channel behaviour. Figure 20 shows the fixed rate and Fig. 21 the Poisson case, respectively. In these figures the mean channel delay is plotted as a function of the channel load, together with some other parameters (such as percentile values, the standard deviation, the maximum and minimum values) in order to give an idea about the delay distribution function. Figure 22 gives a complete example of the delay distribution function, relevant to the 85% channel loading run, already shown in Fig. 20.

The original goal was to make all the stations experience the same delay, independently of the load level. This is almost achieved especially in the Poisson case (Fig. 19), where the saturation effect is evident with the 85% load. Only station 1 is penalised due to the adoption of the allocation limit w_{max} . Figures 20 and 21 show how the Poisson traffic penalises the system delays. Apart from the run in which station 1 is saturated, the mean value of the delay is not dramatically higher in the Poisson case than in the fixed rate case (305÷340 ms versus 295÷315 ms). On the other hand, the difference is noticeable in the spread of values (95^{o} percentile and particularly the maximum values).

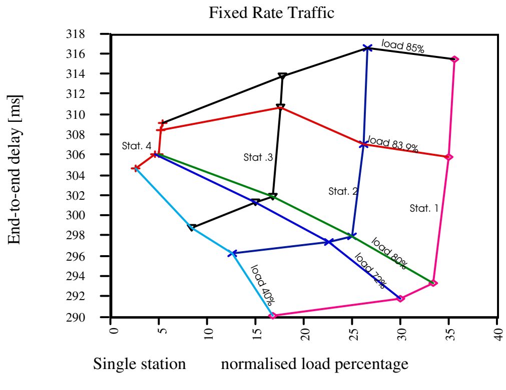

Fig. 18. Single station mean delay versus station load for fixed rate traffic.

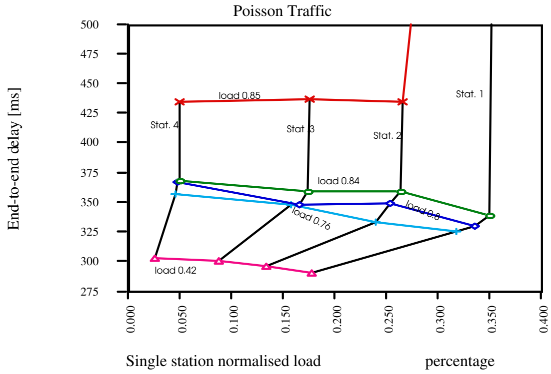

Fig. 19. Single station mean delay versus station load for Poisson traffic.

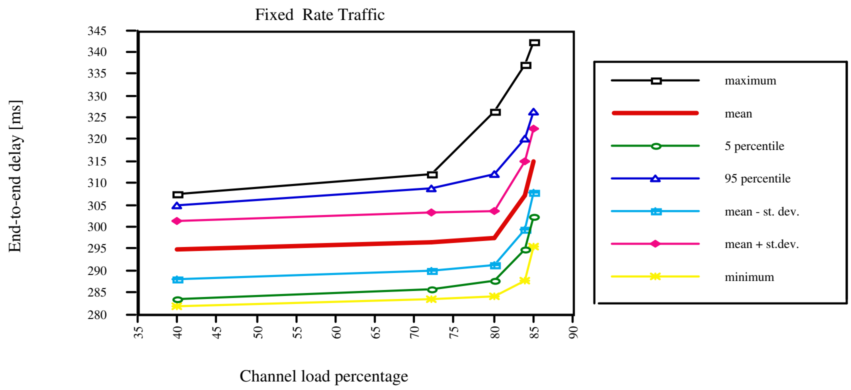

Fig. 20. Channel end-to-end delay statistics versus the channel load for fixed rate traffic.

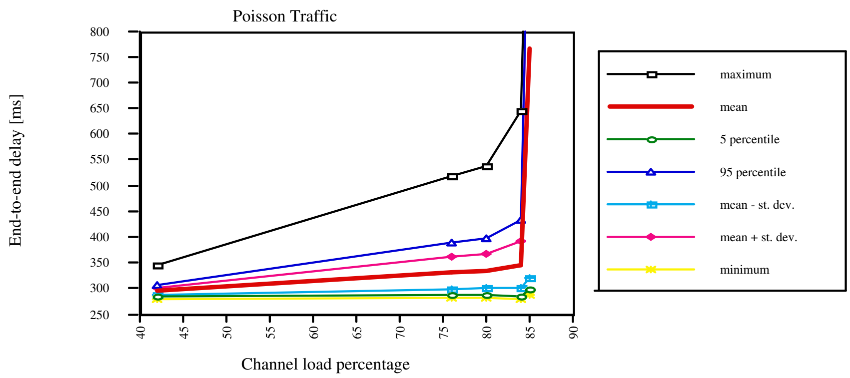

Fig. 21. Channel end-to-end delay statistics versus the channel load for Poisson traffic.

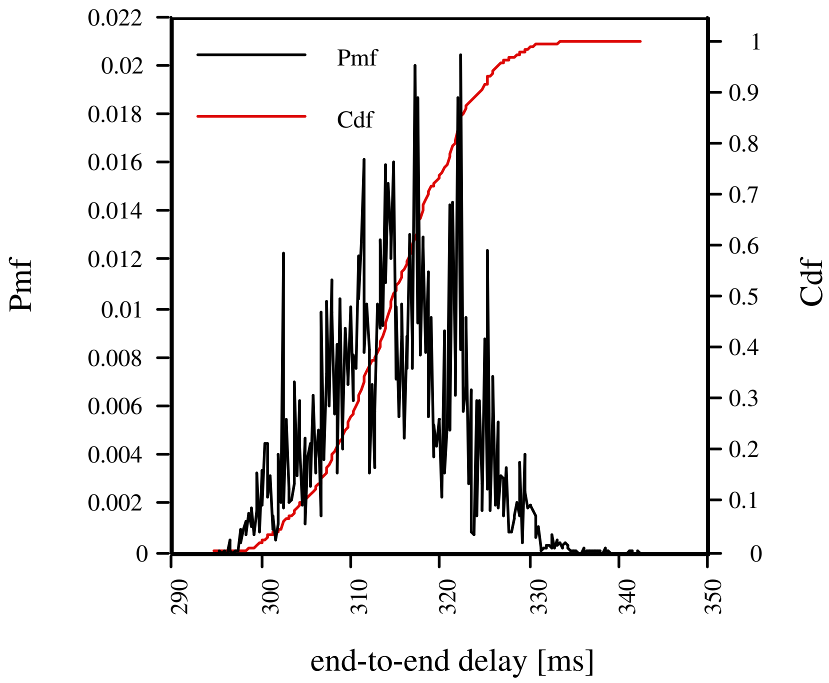

Fig. 22. End-to-end delay distributions: probability mass function and cumulative distribution function. Channel loaded with 85% fixed rate traffic.

3.2.2 The transient case

The transient case was only tested with fixed rate traffic. A traffic step lasting about 6 s fed station number 1 in all three cases presented, which differ from each other in the load levels before the traffic step and in the amplitude of the step itself.

While in the run in Figure 23 the traffic step pushes the system into saturation, in the runs in Figures 24 and 25 this traffic step does not saturate the system which is able to absorb, after a few seconds, the effect due to the increase in traffic. Before the traffic step, the system is scarcely loaded (64% in Figures 23 and 24, 48% in Figure 26). Each station receives an over allocation (pre-assignment) which can absorb, without any visible change, any traffic step lower or equal to the over assigned capacity. This case is obviously not represented. In the cases presented, after the beginning of the traffic step there is an increase in the queue length of the packets to send and, consequently, in the end-to-end delay, due to the system response.

The over assigned capacity is immediately used by station 1 which absorbs part of the traffic step. After a delay of more than two round trips, a bigger allocation is received from the master station and the maximum possible allocation ($w_{\max}$) is assigned to station 1. This is sufficient to curb queue growth and delay. Afterwards, in the case shown in Figure 23, all the other stations see their queues growing and ask for more allocation. This causes a smaller capacity assignment to station 1, thus provoking a new increase in the queue until the end of the traffic step. The transmission window assigned to station 1 is equal to the maximum threshold but it is still insufficient to entirely compensate for the traffic step. This is why station 1 goes into saturation, thus affecting, within limits, the other three stations for the entire duration of the traffic step.

In the case shown in Figure 24, the traffic step of station 1 is not so hard as to saturate the system. So, the larger allocation to station 1 temporarily involves the other stations (but only for the time required by station 1 to absorb the queue caused by the system response delay), whose end-to-end delay briefly increases. The transient in station 1 is also extinguished quite quickly. Figure 25 shows the bursty queue of the four stations relevant to this case.

In Figure 26, stations 2, 3 and 4 practically have no knowledge of the traffic step at station 1. The capacity assigned to station 1 is in any case limited by the maximum threshold and is lower than the one requested. This is why the delay decreases more slowly than in the case shown in Figure 24.

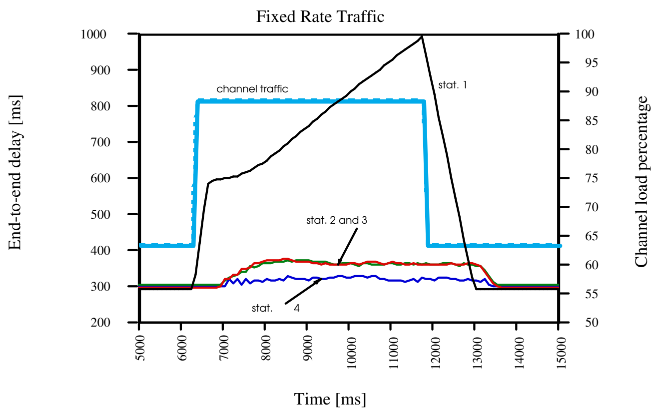

Fig. 23. Single station delay versus time when station 1 has a traffic step of 25% *of the channel capacity (step duration = 5.5 s).*

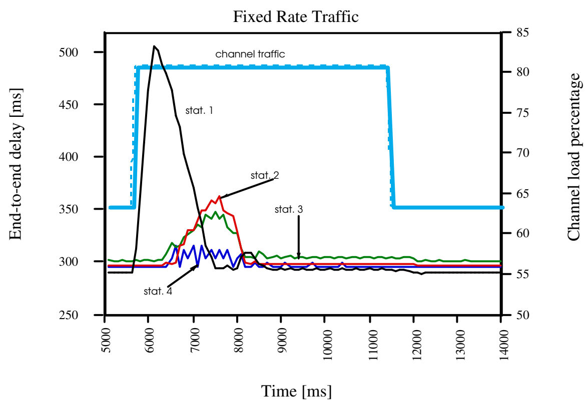

Fig. 24. Single station delay versus time when station 1 has a traffic step of 17% of the channel capacity (step duration = 5.5 s).

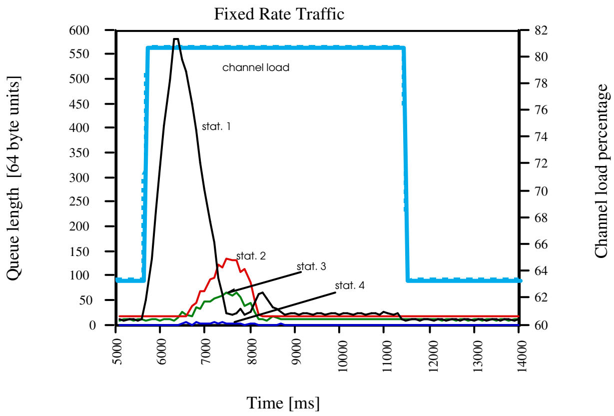

Fig. 25. Queue length versus time when station 1 has a traffic step of 17% of the channel capacity (step duration = 5.5 s).

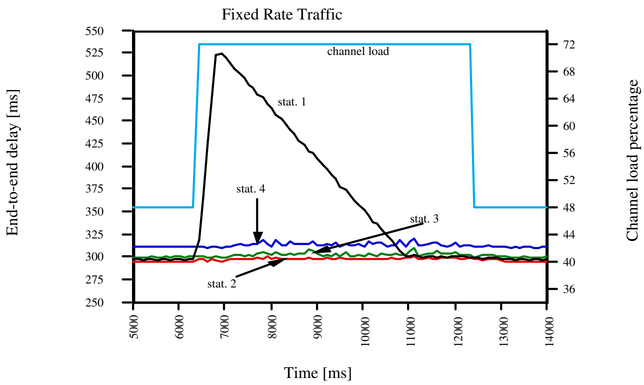

Fig. 26. Single station delay versus time when station 1 has a traffic step of 24% of the channel capacity (step duration = 5.5 s).

4. CONCLUSIONS

We have presented the experimental results relevant to the jitter introduced on the real-time data and the end-to-end delay affecting the non real-time data, when the FODA/IBEA scheme is used to access the satellite link. The analytical results give an indication for the evaluation of the jitter span. In particular, the worst case situation is outlined, in which the packets enter the system already affected by an initial jitter and the system is loaded with all possible traffic patterns. The results measured match the theoretical indications. To remove the jitter we use a leaky bucket. The criteria to size the buffer and the delay offset are also suggested. Sizing the parameters according to the worst case may penalise system performance, thus increasing the end-to-end delay above an acceptable level. This may occur when the initial jitter span is very long, for example if the data arrive at the system after crossing a considerable number of networks. In these cases, in fact, the jitter distribution may have very long tails that contribute negligibly to the cumulative distribution function. In these cases a sub-optimal dimension of the delay offset and of the buffer may be preferable.

As far as the non real-time data is concerned, the stationary cases presented show a limited average end-to-end delay of the data transferred by the satellite network, even in heavy loading conditions. The spread of the delay values is acceptable even when the system is loaded with Poisson traffic generators. The important effect of the maximum allocation threshold is shown when the system is overloaded. In fact, only the most loaded station is heavy penalised in this case.
In the transient analysis, too, the effect of the maximum allocation threshold is outlined, showing the modest perturbation affecting the other stations even when a traffic step of considerable amplitude is applied to a station. The transient may or may not be absorbed by the system but any congestion is limited to the overloaded station.
The convenience of working in pre-assignment mode has been shown by taking the results of the system linear analysis carried out in a referenced paper.

REFERENCES

[1] Horst D. Clausen, Bernard Nocker
"Traffic Characteristics and Resource Control", proceedings of the COST-226 Integrated/Terrestrial Networks, Final Symposium, Budapest, Hungary, 10-12 May 1995.

[2] Francesco Carassa
"Adaptive Methods for 20/30 GHz", Space Communication and Broadcasting 2 (1984), pp. 253-269.

[3] Carassa F., Matricciani E., Tartara G.
"Frequency diversity and its applications", International Journal of Satellite Communications, Vol. 6 N. 3, pg. 313-322, July-September 1988.

[4] Nadery, Kelly
"NASA's advanced communication technology satellite: an overview of the satellite, the network and the underlying techniques", AIAA 12th Communications Satellite Sytems Conference, Arlington, March 1988, pp. 204-224.

[5] M. Plecity, F. Gargione
"Advanced Communications Technology Satellite (ACTS) Applications and Experiments", ICDS-10, Brighton (U.K.) 15-19 May 1995, Vol. 1, pp. 159-166.

[6] Celandroni N., Ferro E., James N., Potortì F.
"FODA/IBEA: a flexible fade countermeasure system in user oriented networks", Inter. Journal of Satellite Communications, Vol. 10, No. 6, pp. 309-323, November-December 1992.

[7] CCITT RED BOOK
Volume III - Fascicle III.3, Digital Networks
Transmission Systems And Multiplexing Equipment.
Recommendations G700 - G956, p. 12

[8] Celandroni N., Ferro E., Potortì F.
"The performance of a TDMA satellite system for non real-time and real time traffic", CNUCE Report C95-12, February 1995.

[9] Wang Zheng, Crowcroft Jon
"Analysis of Burstiness and Jitter in Real-Time Communications", proceedings of the SIGCOMM'93, Ithaca, N.Y., USA, September 1993.

[10] Celandroni Nedo
"Linear analysis of the transient for a bursty demand assignment satellite access scheme in TDMA", CNUCE Report C94-29, December 1994.

[11] Celandroni N. , Ferro E.
"The FODA-TDMA satellite access scheme: presentation, study of the system and results", IEEE Transactions on Communications, Vol. 39, No. 12, pp. 1823-1831, December 1991.

[12] N. Celandroni, E. Ferro, F. Potortì, A. Bellini, F. Pirri
"Practical Experiences in Interconnecting LANs via Satellite", to appear on ACM SIGCOMM Computer Communication Review, October 1995.

[13] Celandroni N., Ferro E.
"The GAFO protocol. The protocol to access the FODA/IBEA system", CNUCE Report C94-03 / Rel.3.0, January 1994.

[14] Celandroni N., Ferro E., Potortì F.
"MTG. Multi-applications Traffic Generator. Presentation and Use", CNUCE Report C95-29 / Rel. 4.0, September 1995.

[15] Harry H. Tan, Santiago V. Hung
"Performance Analysis of slotted ALOHA for LAND Mobile Satellite Networks", Inter. Journal of Satellite Communications, Vol. 10, No. 1, January-February 1992

[16] T-Y. Yan and L.P. Clare
"Performance Analysis of replication ALOHA for fading mobile communications channels", IEEE Trans. on Communications, COM34, (12), 1256,-1266 (1986).

[17] Suzuki M., Kashio J., Toki R., Kato T.
"An Adaptive Control On Length Of Jitter Absorbing Buffer In Packet Speech System", proceedings of the ICC'81, Vol. 2, ref. 41.6, 14-18 June 1981, Denver, Colorado (USA).

[18] Peter David W., Murthy K.M.S., Frost Victor S., Neir Lyn A.
"Modelling and Simulation of the Resource Allocation Process in a Bandwidth-on-demand Satellite Communication Network", IEEE Journal on Selected Areas in Communications, Vol. 10, N.2, February 1992, pp. 465-477.

[19] Chang Jin-Fu, Lin Sung-Hui
"Delay Performance of VSAT-Based Satellite Wide Area Networks", International Journal of Satellite Communications, Vol. 11, No. 1, January-February 1993, pp.1-12.

[20] Tasaka Shuji, Ishibashi Yutaka

"A Reservation Protocol for Satellite Packet Communication. - A Performance Analysis and Stability Considerations", IEEE Transactions On Communications, Vol. Com-32, N. 8, August 1984, pp. 920-927.

[21] Priscoli Delli F., Listanti M., Roveri A., Vernucci A.
"A Distributed Access Protocol For An ATM User Oriented Satellite System", ICC '89, Boston (USA), June 11-14, Vol. 2, pp. 690-694.

[22] Lam S.S.
"Satellite Packet Communication. Multiple Access protocol and Performance", IEEE Transactions on Communications, Vol. COM-28, No. 4, April 1980, pp. 468-488.

[23] Papantoni-Kazakos P., Hall Thornton
"Multiple Access Algorithms For a System With Mixed Traffic: High and Low Priority", ICC '89, Boston (USA), June 11-14, Vol. 2, pp. 665-671.