5.5.4 Bandwidth Assignment Strategies and Analytical Modeling of Elastic Traffic for Dynamic Resource Allocation in Multiservice Satellite Networks with Fading

Research Unit: CNIT - University of Genoa (DIST section)

Introduction

In multi-service broadband networks, in order to support different classes of traffic, characterized by diverse statistical nature and Quality of Service (QoS) requirements, in the presence of limited resources (buffers, bandwidth, or processing capacity), several forms of control are exerted to maintain a desired level of performance for all users and traffic types. Moreover, to cope with possible variations in bandwidth demand and offered load, control actions should be devised to be dynamic, based on instantaneous and past information, or, at least, adaptive in nature. This control scenario meets with still another one in networks involving satellite channels. Here, dynamically varying fading conditions, caused by adverse atmospheric events (e.g., rain, hail or snow) can heavily affect the transmission quality, especially when working in Ka band [20 – 30 GHz] [1-3], unless adaptive fade countermeasure techniques are adopted. In essence, fade countermeasures address the physical layer requirement of keeping the Bit Error Rate (BER) below a given threshold, whatever the channel degradation may be, within a certain operating range, beyond which the station is declared to be in outage conditions.

On the other hand, controls applied to cope with variations in traffic dynamics and with the presence of multiple services, in order to guarantee potentially different performance requirements and to avoid congestion, typically act at protocol layers from 2 above, spanning data link, MAC, network, transport, and even the application layers. In many cases, the choice of where, how and for what purpose to exert a control action depends on the type of traffic to be dealt with and on the extent of service separation, as opposed to complete statistical multiplexing, that is adopted for packetized traffic. In particular, in the presence of both guaranteed bandwidth and best-effort flows, either deriving from elastic traffic, like TCP connections, or from non-guaranteed real-time applications using UDP, the probability of blocking connection requests and the loss of data units in buffers are two important performance indexes.

In this work, two important aspects have been taken in account:

- dynamic bandwidth allocation strategies to compensate the rain attenuation of the satellite channel;
- allocation strategies based on analytical traffic models for elastic traffic.

Dynamic Resource Allocation Strategies

Several methods have been studied in the literature and effectively applied to provide compensation for rain attenuation; among others, we recall using additional transmission power (up-power control), and using a portion of the system bandwidth to have a rain margin. The latter may be used to increase the data coding rate for stations that temporarily suffer heavier attenuation, by adopting appropriate Forward Error Correction (FEC) codes in order to maintain the desired BER level, even by decreasing the information transmission rate, if necessary, when no more spare bandwidth is available. In general, we have a resource allocation problem, consisting of the assignment of a total available bandwidth among the earth stations and, inside them, among the traffic types. We consider some possible approaches that essentially combine the physical layer actions related to fade countermeasures with CAC and bandwidth allocation. The emphasis is on a decentralized administration of the assigned bandwidth, with some kind of hierarchical coordination. Satellite networks almost naturally lend themselves to this kind of approach, as the earth stations are scattered on a wide area and, at the same time, a master station, playing the role of a coordinator, is usually present. We distinguish two situations, as regards the information available on-line to the master station on the level of fading attenuation of the traffic stations. According to this, the assignment can be made static, i.e. based on the a priori knowledge of long-term fading statistics, or dynamic, based on fresh values of the attenuation levels of the traffic stations. In either case, it is possible to adapt the assignment on-line to slowly time-varying traffic characteristics. At each earth station, two basic traffic types are supposed to be present: i) guaranteed bandwidth, real-time, synchronous traffic (stream traffic), generated by voice and video applications, and ii) non-guaranteed bandwidth, non real-time, asynchronous traffic (datagram traffic) [4-6], produced by data applications.

In dynamic allocation schemes, an important problem should be taken in account: the reallocation timing. Three different strategies, based on traffic load evolution and channel state variation, are considered: in the first one the bandwidth reallocation is performed synchronously, at the beginning of each fixed length data frame, while in the second one the bandwidth is reallocated on request of any station that cannot afford a fade level variation or an increment in the average traffic offered. The request is triggered by insufficient capacity to continue carrying all connections in progress. The third strategy considers both events for making a reallocation. The three different strategies are compared in terms of call blocking and cell loss probabilities for 10 stations and for the three different strategies, by using several attenuation data, taken from a data set chosen from the results of the propagation experiment, in Ka band, carried out on the Olympus satellite by the CSTS (Centro Studi sulle Telecomunicazioni Spaziali) Institute, on behalf of the Italian Space Agency (ASI). The up-link (30 GHz) and down-link (20 GHz) samples considered were 1-second averages, expressed in dB, of the signal power attenuation with respect to clear sky conditions. The attenuation samples were recorded at the Spino d’Adda (North of Italy) station, in September 1992.

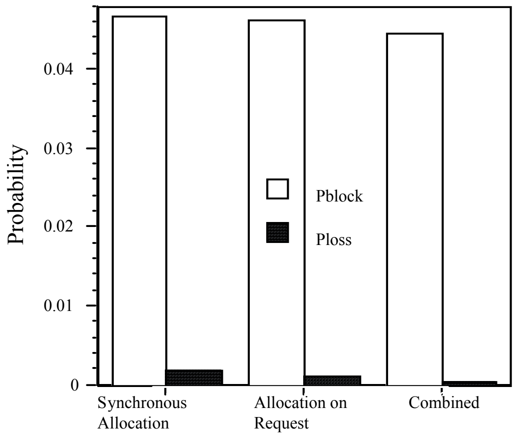

Fig. 1: Call blocking and Data Unit loss probabilities of the entire system for the three strategies.

The synchronous allocation is not able to follow the rapid fade changes, thus giving rise to non-negligible periods where no capacity is available for datagram. On the other hand, the reallocation on request, used alone, does not release the excess of bandwidth obtained to overcome the fade events, thus starving the other stations. The adoption of both methods, instead, solves all drawbacks and gives appreciable results, where the call blocking and data loss probabilities of the entire system, averaged over all stations, are lower in the case of the combined strategy. The overall performance of the strategies proposed is shown in figure 1 below.

The work done opens the way for further interesting investigations. An interesting point regards the distribution of the optimization task among the traffic stations or, at the opposite extreme, the centralization of the CAC for the whole system at the master station, in a "complete sharing" fashion. In fact, there is a whole range of possible control architectures that can be investigated, and the one considered here is an intermediate case of hierarchical coordination. In particular, it is worth noting that there is a form of looser coordination, which gives more autonomy to the local controllers, though still keeping a role for the master in the bandwidth assignment. In this case, the earth stations, rather than communicating their fade and traffic parameters to the master, independently compute the bandwidth they would need to satisfy their requirements.

Bandwidth Allocation With Elastic Traffic

Most of the applications currently on the market are TCP/IP-based. On the other hand, satellites offer clear advantages with respect to cable networks: scalability, wide diffusion through the land, overcoming of geographical obstacles. Matching the applications that use TCP/IP with the advantages offered by satellites, it is natural to think of TCP/IP-based applications over satellite networks. The satellite network considered is composed of earth stations connected through a geostationary satellite. An earth station or the satellite itself represents the master, which manages the resources and provides the other stations with a portion of the overall bandwidth; each of the latter shares the assigned portion between the traffic flows of the single station. The fading is modeled by assigning a probability of channel degradation with a weighting coefficient to 'measure' the degradation itself. Three types of traffic are considered: a QoS guaranteed traffic, modeled as synchronous transmission operating at a fixed speed and two non-guaranteed best-effort traffics: UDP, modeled by a self-similar Pareto distribution [6], and TCP, whose packet loss models have been introduced.

The analytical expression of the TCP connection packet loss probability introduced during the project is a function of the bandwidth available and it has been employed in the bandwidth allocation schemes proposed for GEO satellites. This work, starting from the state-of-the-art [7-9], proposes a closed formula of the TCP packet loss probability (p_n), in dependence of the overall number of active connections (N) of the round trip time imposed by the satellite channel (RTT) for each *n*-th TCP connection, of the buffer (Q) and of the bandwidth available on the channel (C):

$$p_n = \frac{6N^2}{b_n \cdot (m+1)^2 \cdot (C \cdot RTT + Q)^2} \tag{1}$$

In equation (1), the other parameters indicated are b_n, the number of packets covered by one acknowledgement and *m*, the congestion window reduction imposed by the TCP protocol during the congestion avoidance phase. The efficiency of the formulation is evaluated through a satellite emulator. The analytical results found with the model, compared with measures obtained by the hardware satellite emulator, have shown a high degree of accuracy.

The TCP packet loss formulation, together with the CBR and UDP models, is used to get a new cost function and a bandwidth allocation scheme (called E-CAP-ABASC, where E stands for Extended), which represents the novelty. The aim of the performance evaluation provided to test the new strategy proposed is to compare the packet loss probability measure used in E-CAP-ABASC with the results obtained with the packet loss probability used in the former CAP-ABASC. To perform the comparison fairly, the two formulas need to be evaluated with the same bandwidth allocated. The tests have been performed with 4 earth stations. Stations 1, 2 and 3 are in clear sky conditions; station 4 varies its condition as indicated below (where β_i is the fading level and P_i its related probability) and the overall bandwidth available is set to 8 [Mbps]. The packet loss estimation shows that considering IP traffic undistinguished gives origin to an overestimation. It is due to the fact that CAP-ABASC ignores the self-reduction of the transmission rate operated by the TCP congestion avoidance, which, on the contrary, is considered by E-CAP-ABASC. The measure of the difference is revealed by figure 2. The difference is less evident in TEST3 and in TEST4, due to the reduced

	β_1	P_1	β_2	P_2
TEST1	1	1	0	0
TEST2	0.6	1	0	0
TEST3	0.4	0.5	0.3	0.5
TEST4	0.3	0.5	0.2	0.5

Table 1: Fading levels and related probabilities.

bandwidth availability imposed by penalizing fading levels. The different estimation may affect bandwidth allocation, allowing to save bandwidth with the same overall performance.

In other words, if a packet loss constraint is imposed, overestimating means more and wasted bandwidth. The control architecture for satellite systems introduced can distinguish TCP from UDP traffic and, in future developments, can "open the door" to different cost functions and related allocation schemes, where TCP/IP traffic and applications play a major role.

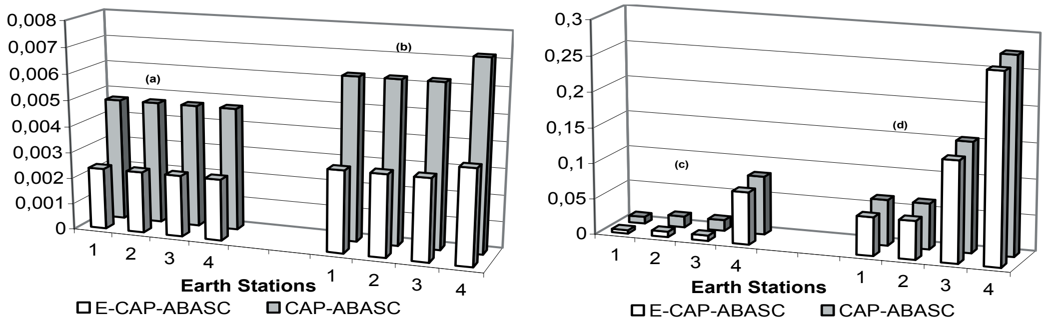

Fig. 2: Packet loss probability comparisons for TEST1 (a), TEST2 (b), TEST3 (c) and TEST4 (d).

References

[1] N. Celandroni, F. Davoli, E. Ferro, S. Vignola, S. Zappatore, A. Zinicola, "Quality of service measurements of video encoded sequences over an emulated Ka band satellite environment", Proc. 8th Ka band Utilisation Conference, Baveno, Italy, Sept. 2002, pp. 429-437.

[2] R Bolla, N. Celandroni, F. Davoli, E. Ferro, M. Marchese, "Bandwidth allocation in a multiservice satellite network based on long-term weather forecast scenarios", Computer Communications, vol. 25, no. 11-12, July 2002, pp. 1037-1046.

[3] N. Celandroni, F. Davoli, E. Ferro, "Static and dynamic resource allocation in a multiservice satellite network with fading", International Journal of Satellite Communications and Networking, Special Issue on Satellite IP Quality of Service, vol. 21, Issue 4-5, pp. 469-487, July-Oct. 2003.

[4] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, "On the self-similar nature of Ethernet traffic", IEEE/ACM Transactions on Networking, 1994, vol. 2, no. 1, pp. 1-15.

[5] M.E. Crovella, A. Bestravos, "Self-similarity in World Wide Web Traffic: Evidence and possible causes", IEEE/ACM Transactions on Networking, 1997, pp. 835-856.

[6] B. Tsybakov, N.D. Georganas, "On self-similar traffic in ATM queues: definition, overflow probability bound, and cell delay distribution", IEEE/ACM Transactions on Networking, 1997, pp. 397-409.

[7] E. Altman, C. Barakat, E. Laborde, P. Brown, D. Collange, "Fairness analysis of TCP/IP", Proc. IEEE Conference on Decision and Control, Dec. 2000.

[8] T. Lakshman, U. Madhow, "The performance of TCP/IP for networks with high bandwidth-delay products and random loss", IEEE/ACM Transactions on Networking, vol. 5, no. 3, 1997.

[9] J. Padhye, V. Firoiu, D. Towsley, J. Kurose. "Modeling TCP throughput: A simple model and its empirical validation", Proc. Sigcomm 1998.