Packet Loss and Delay Combined Optimization for Satellite Channel Bandwidth Allocation Controls

Igor Bisio *Member, IEEE*, Mario Marchese *Senior Member, IEEE*
DIST - Department of Communication, Computer and System Science
University of Genoa, Via Opera Pia 13, 16145, Genoa, Italy
phone: +39-010-3532806, fax: +39-010-3532154
e-mails: { igor.bisio, mario.marchese }@unige.it

***Abstract*—The paper studies the bandwidth allocation process over satellite communication systems as a Multi – Objective Programming (MOP) problem and evaluates an allocation method called "Combined Utopia Minimum Distance" (Combined UMD). The entities of the system are earth stations and, for each of them, a set of performance metrics (represented by specific analytical functions), which compete to access the satellite channel. Combined UMD is aimed at approaching the performance obtained for each performance metric when there is no conflict among them to access the channel. In short, it assigns the bandwidth so to approach a non-competitive situation where each metric "sees" the overall channel bandwidth availability as close as possible. In more detail, in this work two kinds of performance metric have been considered: the Packet Loss Probability, which is a typical QoS metric for the TCP based traffic and the Average Delay, which is typical for UDP based traffic. The allocation method is tested through the *ns2* simulator by using TCP and UDP traffic generators and by varying the fading level of the satellite channel over time. Combined UMD has been compared with other approaches taken from the literature in the field.**

Keywords-Satellite Systems, QoS Metrics, Multi-Objective Programming, Bandwidth Allocation, Performance Evaluation.

I. INTRODUCTION

In Satellite environment [1] bit errors caused by noise and atmospheric conditions (e.g., rain fading) are major issue and object of investigation. High Bit Error Rates (BERs) affect the ability of the satellite channel to offer reliable data transmission. Forward Error Correction (FEC) Coding, typically employed in these systems, is able to compensate for such errors, trading bandwidth for effective data (as shown in [2, 3] and synthetically reported in Section II.A.). On the other hand, Quality of Service (QoS) is the ability of a network element (e.g. an application, host, router or, in this case, satellite gateway) to have some level of assurance so that its traffic and service requirements can be satisfied. In the considered environment communication detriment due to BER and Fading causes QoS degradation.

Allocating the bandwidth properly among satellite earth stations, which can be affected by different noise or fading levels, is important to mitigate the problem and to increase the provided QoS level. The rationale under this paper is considering bandwidth allocation as a competitive problem, by extending the concept proposed in [2], where each station is "represented" by a set of performance functions, each of them representative of a traffic with different QoS sensitivity, which needs to be minimized at cost of the others. All functions must be minimized simultaneously. In more detail, in this work, data traffic has been categorized into two major types with different QoS sensitivity: delay sensitive traffic (real-time) and loss sensitive traffic (non-real-time) traffic. Delay Sensitive applications are multimedia applications such as video conferencing, and Internet telephony (VoIP) typically based on the UDP transport layer protocol. Loss sensitive applications are those not involving multimedia data, such as FTP, that use the TCP transport layer protocol.

In practice, stations, and for each of them, traffics with different QoS requirements compete for bandwidth. It is the definition of the Multi-Objective Programming (MOP) class of problems, which is the base of the method employed in the paper. Combined Utopia Minimum Distance (Combined UMD) is aimed at approaching the ideal performance, which theoretically happens when each single performance function of an earth station is considered alone and has the availability of all channel bandwidth.

Combined UMD approach allows, differently from other methodologies in the literature, the combined optimization among competitive and heterogeneous QoS metrics. As previously mentioned, two traffic classes have been considered: Loss Sensitive Traffic and Delay Sensitive Traffic. Each of them is supposed to have a dedicated buffer. In the heterogeneous QoS metric sensitivity case considered, previously proposed approaches in the literature, such as the optimization of the sum of the performance functions, would not work well. For example, if the buffers have equal and small sizes the packet loss probability of the loss sensitive traffic is very high. Vice versa, being low the number of packets in the buffers the delay (seen as queue waiting time) of the delay sensitive traffic is very low. If the optimization approach, used to allocate the bandwidth among stations, is based on the sum of all metrics the delay might not play any role. The MOP approach solve this possible inconvenient.

The paper is structured as follows: Section II introduces the Network Structure and the control architecture. The formalization of the bandwidth allocation as a MOP problem and the Combined Utopia Minimum Distance criterion is presented in Section III. Section IV reports an introductive performance evaluation obtained through the *ns2* simulator. Section V lists the conclusions.

978-1-4244-2075-9/08/$25.00 ©2008 IEEE

II. NETWORK STRUCTURE AND CONTROL ARCHITECTURE

A. *Geostationary Satellite Network*

The network considered is composed of earth stations connected through a satellite link. Each user requests a TCP/IP service (e.g., Web page, File transfer or a VoIP session) by using the satellite channel itself (or also other communication media). After receiving the request ISPs send traffic through the earth stations and the satellite link. To carry out the process, each earth station conveys traffic from the directly connected ISPs and accesses the channel in competition with the other earth stations.

In the considered satellite network, as mentioned in the introduction of the work, noise and fading effects, which may affect the QoS performance of the network, are modelled as bandwidth reduction. From the practical viewpoint, it means using a FEC code where each earth station may adaptively change the amount of redundancy bits (e.g. the correction power of the code) in dependence on noise or fading, so reducing the real bandwidth availability. Mathematically, it means that the bandwidth $C_z^{real} \in \mathbb{R}$ available for the z-th station is composed of the nominal bandwidth $C_z \in \mathbb{R}$ and of the factor $\beta_z \in \mathbb{R}$, which is, in this paper, a variable parameter contained in the interval [0, 1]. Formally, as in [2, 3], $C_z^{real} = \beta_z \cdot C_z$. A specific value β_z corresponds to a fixed attenuation level "seen" by the z-th station. An example of the mapping between the Carrier Power to One-Side Noise Spectral Density Ratio (C/N_0) and the β_z parameter is contained in reference [3].

B. *Control Architecture*

The control architecture is based on the presence of decision entities, also called Decision Makers (DMs) as reported in Fig. 1. In general, it can be used one DM for the whole system in a centralized way, where an earth station or the satellite itself, if switching on board is allowed, represents the single DM that manages and provides stations with a portion of the overall bandwidth (e.g., TDMA slots). Alternatively, may be used a distributed implementation of the bandwidth allocation process where one DM for each station is employed. In this work, the second approach has been provided: each station has a DM that manages the bandwidth distribution independently of each others (Fig. 2). From the structural viewpoint, each station has a battery of buffer, one for each kind of traffic. In this case, one buffer is dedicated to Loss Sensitive Traffic (TCP based) and one dedicated to Delay Sensitive Traffic Source (UDP based). The first buffer is employed to store and, consequently send, the TCP packet (e.g., traffic generated from TCP based sources such as FTP sessions or Web Browsing applications); the second buffer is used by UDP traffic (e.g., traffic generated from UDP based sources such as VoIP sessions or Video Streaming applications). Each buffer of a station has a service capacity that is a portion of the capacity allocated, by the DM(s) to the station itself.

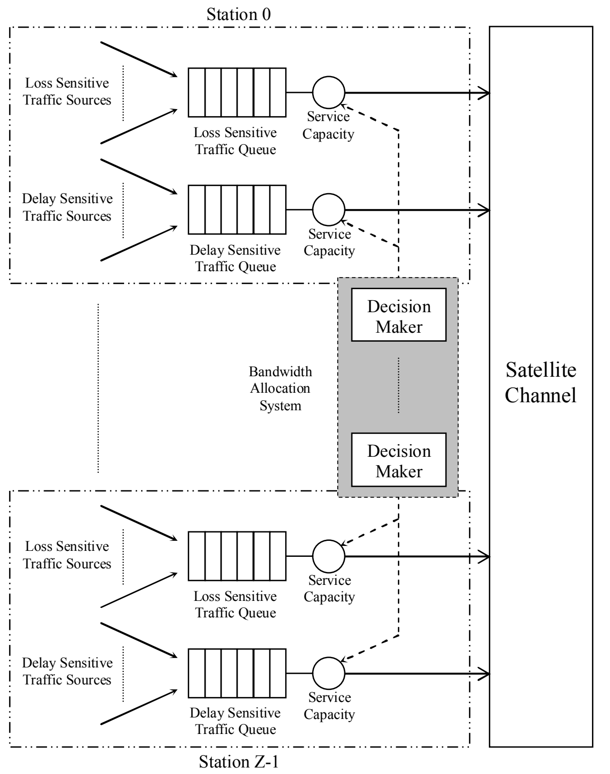

Fig. 1. Control Architecture.

III. LOSS AND DELAY COMBINED OPTIMIZATION

A. *Multiple QoS Metrics Problem Formulation (MOP)*

The practical aim of the allocator is the provision of bandwidth to each k-th $\forall k \in [0, K-1], k \in \mathbb{N}$ buffer server of each z-th station $\forall z \in [0, Z-1], z \in \mathbb{N}$ by splitting the overall capacity available among the buffers (the competitive entities of the problem). Each buffer is employed to serve a specific kind of traffic and, as a consequence, it is represented by a specific performance metric. Analytically, the bandwidth allocation defined as a Multi – Objective Programming (MOP) problem may be formalized as:

$$\begin{aligned}\mathbf{C}^{opt} &= \left\{C^{opt}_{0,0},...C^{opt}_{0,K-1};...C^{opt}_{z,0},...C^{opt}_{z,K-1};...C^{opt}_{Z-1,0},...C^{opt}_{Z-1,K-1}\right\} \\ &= \arg\min_{\mathbf{C}}\left\{\mathbf{F}(\mathbf{C})\right\};\ \mathbf{F}(\mathbf{C}) : \mathbf{D} \subset \mathbb{R}^{Z\times K} \to \mathbb{R}^{Z\times K}\end{aligned} \quad (1)$$

where: $\mathbf{C} = \left\{C_{0,0},...,C_{0,K-1};...,C_{z,0},...,C_{z,K-1};...,C_{Z-1,0},...,C_{Z-1,K-1}\right\}$, $\mathbf{C} \in \mathbf{D}$, is the vector of the capacities assignable to the earth stations' buffers; the element $C_{z,k}$, $\forall z \in [0,\ Z-1], \forall k \in [0, K-1], z, k \in \mathbb{N}$ is referred to the k-th buffer of the z-th station; $\mathbf{C}^{opt} \in \mathbf{D}$, is the vector of the optimal allocation; and $\mathbf{D} \subset \mathbb{R}^{Z\times K}$ represents the domain of the vector of functions. The solution has to respect the constraint:

$$\sum_{z=0}^{Z-1}\sum_{k=0}^{K-1} C_{z,k} = C_{tot} \tag{2}$$

where C_{tot} is the available overall capacity. $\mathbf{F}(\mathbf{C})$, dependent on the vector $\mathbf{C}$, is the performance vector

$$\begin{aligned}\mathbf{F}(\mathbf{C}) = \{ & f_{0,0}(C_{0,0}),\ldots f_{0,K-1}(C_{0,K-1});\ldots \\ & \ldots f_{z,0}(C_{z,0}),\ldots f_{z,K-1}(C_{z,K-1});\ldots \\ & \ldots f_{Z-1,K-1}(C_{Z-1,K-1}),\ldots f_{Z-1,K-1}(C_{Z-1,K-1})\}\end{aligned} \tag{3}$$

The single z,k-th $\forall z \in [0, Z-1], \forall k \in [0, K-1]\ Z, K \in \mathbb{N}$ performance function is a component of the vector. Each performance function $f_{z,k}(C_{z,k})$ (or objective) of the system is defined here as the average packet loss probability for the TCP traffic (if $k=0$) and as the average delay for the UDP traffic (if $k=1$). As a consequence, $K=2$ possible performance metrics have been considered. Actually any other convex and decreasing function of the bandwidth may be used.

The packet loss probability for TCP traffic and the delay for UDP seems a reasonable choice but it may be regarded also as an operative example for the theory presented.

B. TCP Packet Loss Probability (PLP)Function

The used TCP packet loss probability $P^z_{loss}(\cdot)$ is a function of the bandwidth ($C_{z,0}$) as well as of the number of TCP active sources (N_z) and of the fading level (β_z), for each station z. $P^z_{loss}(\cdot) = f_{z,0}(\cdot)$ may be expressed as:

$$\begin{aligned} & P^z_{loss}(C_{z,0}, N_z, \beta_z) = \\ & = 32 N_z^{\,2} \cdot \left[3b(r+1)^2 \left(\beta_z \cdot \widetilde{C_{z,0}} \cdot RTT + \widetilde{Q_{z,0}} \right)^2 \right]^{-1} \end{aligned} \tag{4}$$

where: b is the number of TCP packets covered by one acknowledgment; r is the reduction factor of the TCP transmission window during the Congestion Avoidance phase (typically $r = 1/2$); $\widetilde{C_{z,0}}$ is the bandwidth "seen" by the TCP aggregate of the z-th earth station expressed in packets/s ($\widetilde{C_{z,0}} = C_{z,0}/d_0$, where d_0 is the TCP packet size, always fixed in this paper); RTT is the Round Trip Time; $\widetilde{Q_{z,0}}$ is the buffer size, expressed in packets, of the z-th earth station dedicated to the TCP traffic.

The used PLP ($P^z_{loss}(\cdot)$) is a monotone decreasing and convex function $\forall C_{z,0} \geq 0,\ \forall z \in [0,\ Z-1],\ z \in \mathbb{N}$, and it considers the effect of the channel state because it is also a function of the β_z parameter. The model is valid at regime condition of the TCP senders.

C. UDP Average Delay (AD) Function

The UDP average delay, defined as the delay spent in the UDP buffer of the z-th earth station, $D^z(\cdot)$ is a function of the bandwidth ($C_{z,1}$) as well as of the number of UDP active sources (M_z) and of the fading level (β_z), for each station z. $D^z(\cdot) = f_{z,1}(\cdot)$ may be expressed taking as reference the well known M/M/1/X model, where X is equal to the overall storage capacity in packets (buffer size plus packet in service), of the UDP buffer. In this case, for each UDP buffer of earth stations $X = \widetilde{Q_{z,1}} + 1$. Starting from the M/M/1/X model hypothesises the average delay is:

$$D^z(C_{z,1}, M_z, \beta_z) = \frac{1}{\mu_z} \cdot \left[\left(\frac{\rho_z}{1-\rho_z} \right) - \left(\frac{\left(\widetilde{Q_{z,1}}+2\right) \rho_z^{\widetilde{Q_{z,1}}+2}}{1-\rho_z^{\widetilde{Q_{z,1}}+2}} \right) \right] \tag{5}$$

where: $\mu_z = (C_{z,1} \cdot \beta_z)/d_1$ is the UDP buffer service capacity of the z-th earth station; d_1 is the UDP packet size, always fixed in this paper; $\rho_z = \lambda_z/\mu_z$ is the offered load to the UDP buffer of the z-th earth station; $\lambda_z = \sum_{m=0}^{M_z-1} \lambda_z^m$ is the overall arrival rate of the UDP packets in its dedicated queue (λ_z^m is the generation rate of the single m-th UDP source); $\widetilde{Q_{z,1}}$ is the UDP buffer size, expressed in packets, of the z-th earth station.

Also the used AD ($D^z(\cdot)$) is a monotone decreasing and convex function $\forall C_{z,1} \geq 0,\ \forall z \in [0,\ Z-1],\ z \in \mathbb{N}$, and it considers the effect of the channel state because it is also a function of the β_z parameter. The model is valid at regime condition of the UDP senders.

It is worth noting that the model proposed, whose computation has been omitted for the sake of synthesis, has been used by relaxing the hypothesis of exponential distribution of the service time. In this paper a deterministic service time has been applied.

D. Multiple QoS Metrics Allocation (Combined UMD)

In general, the problem defined above, is a Multi – Object Programming problem where each considered function $f_{z,k}(C_{z,k})$ represents a single competitive cost function. In other words, a single performance function competes with the others for bandwidth. The optimal solution for MOP problems is called POP-Pareto Optimal Point, coherently with the classical MOP theory.

The Utopia Minimum Distance method is a flexible methodology that allows the resolution of the allocation problem (1). It bases its decision only on the ideal solution of the problem: the so called utopia point. In more detail, the ideal performance vector, in the case of this work, is:

$$\begin{aligned}\mathbf{F}^{id}\left(\mathbf{C}^{id}\right)=\Big\{&f_{0,0}^{id}\left(C_{0,0}^{id}\right),\dots f_{0,K-1}^{id}\left(C_{0,K-1}^{id}\right);\dots\\ &\dots f_{z,0}^{id}\left(C_{z,0}^{id}\right),\dots f_{z,K-1}^{id}\left(C_{z,K-1}^{id}\right);\dots\\ &\dots f_{Z-1,0}^{id}\left(C_{Z-1,0}^{id}\right),\dots f_{Z-1,K-1}^{id}\left(C_{Z-1,K-1}^{id}\right)\Big\}\end{aligned} \tag{6}$$

where

$$f_{z,k}^{id}\left(C_{z,k}^{id}\right)=\min_{C_{z,k}}\left[f_{z,k}\left(C_{z,k}\right)\right],\ C_{z,k}\in\left[0,C_{tot}\right] \tag{7}$$

From equation (7), called single objective problem, it is clear that the optimal solution is given by $C_{z,k}=C_{tot}$, $\forall z\in\left[0,Z-1\right],\forall k\in\left[0,K-1\right]$.

So, $\mathbf{C}^{id}=\left\{C_{tot},C_{tot},\dots,C_{tot}\right\}$. Obviously it is a physically unfeasible condition that can be only approached due to constraint (2). Starting from the definition of the ideal performance vector, the problem in equation (1) can be solved with the following allocation (coherently with [2]):

$$\mathbf{C}^{opt}=\arg\min_{\mathbf{C}}\left(\left\|\mathbf{F}\left(\mathbf{C}\right)-\mathbf{F}^{id}\left(\mathbf{C}^{id}\right)\right\|_{2}\right)^{2} \tag{8}$$

where $\left\|\cdot\right\|_{2}$ is the Euclidean norm. The proposed technique allows minimizing the distance between the performance vector and the ideal solution of the problem. Obviously, the minimization is carried out under the constraint (2).

IV. PERFORMANCE EVALUATION

The aim of this performance evaluation is to evaluate the bandwidth allocation method functionalities in terms of PLP and AD. The action is fulfilled by using an *ns2* based simulator, where the optimization procedures have been implemented. In the following tests, the comparisons have got by varying the fading conditions, in practice a given behaviour of the β_z parameter over time has been used in the simulations for each earth station considered.

The allocator acts periodically (each T_a [s]). In each allocation instant, each DM (Fig. 2) knows the fading level and the traffic parameters, related to its earth station, through a specific signalling procedure. After that, DMs provide the bandwidth allocation, by solving equation (8), in a negligible computation time T_c ($T_c \ll T_a$).

The network scenario considered is composed of $Z=4$ earth stations: Stations from 0 to 2 are always in clear sky condition (β_z always equal to 1 $\forall z\in\left[0,2\right]$), Station 3 varies its fading level, according to real fading levels taken from [3], over time as made explicit in the following figure:

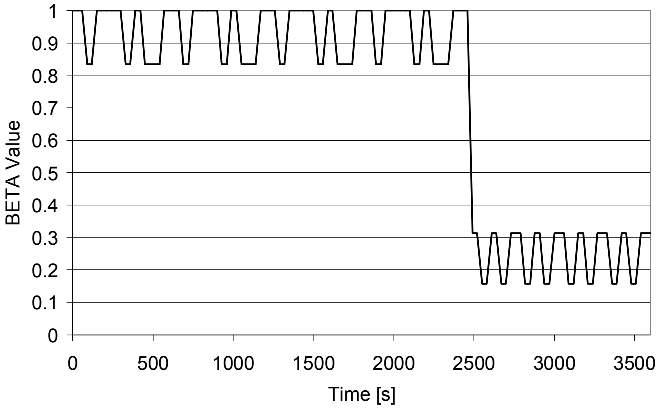

Fig. 2. Fading Level Variation.

Each station gathers traffic from TCP and UDP sources and transmits it to the terminal users through the Satellite system. The number of active TCP and UDP sources is set to $N_z=M_z=5$, $\forall z\in\left[0,Z-1\right]$. The overall bandwidth available C_{tot} is set to 8 [Mb/s] and the TCP and UDP buffer size $\widetilde{Q_{z,0}}=\widetilde{Q_{z,1}}$ is set to 10 packets (of $d_0=d_1=1500$ bytes) for each earth station. The Round Trip Time (RTT) value employed in the computation of the traffic model is supposed fixed and equal to 512 [ms] for all the stations. The allocation control acts each $T_a=30$ [s] and in all cases the simulated time is always fixed and equal to 3600 [s]. The TCP sources actives a FTP session at the beginning of the simulations. Each FTP transfer has been set as a persistent session for the overall duration of the simulations: in practice, sources have always packets to send. The UDP sources are considered Poissonian packet generator with fixed λ_z^m equal to 100 [packets/s]. Each buffer is implemented as a Dumbbell topology with a single common receiving node. The topology is composed of 20 source agents (which are nodes): 10 agents active 1 TCP connection and the others 1 UDP connection. They send their packets to earth stations by using not congested and wideband links that do not represent bottlenecks during simulations. An earth station is, in practice, a pair of buffers with storing capacities equal to $\widetilde{Q_{z0}}$ and $\widetilde{Q_{z1}}$ packet, where packets sent from sources are conveyed and forwarded if no congestion events are experienced. The service capacity, in [b/s], of the buffers of an earth station, is the bandwidth allocated to it and the effect of the fading is considered by using the model mentioned in Section II.A: the fading is supposed completely compensated by using FEC schemes (no channel errors are considered in the simulations) and their impact is a mere bandwidth reduction represented by the β_z parameter.

In Figs. 3 and 4, the measured Packet Loss Probability has been reported. In Fig. 3 a clear sky station has been considered (more specifically Station 0); in Fig. 4 the PLP performance of the faded station (Station 3) has been reported. The Combined UMD technique described in previous sections has been compared with a simple STATIC approach (each buffer of the earth station receive a fixed quantity of bandwidth equal to $1/Z\times K$ of the overall available capacity) and with a method, taken from the literature [4] here called VALUE, where the allocation is obtained by minimizing the sum of the performance functions.

The three techniques have similar PLP performance if the earth stations "see" good channel conditions ($0 < t \le 2500$). When the fading becomes severe ($2500 < t \le 3600$) VALUE and UMD obviously have better performance with respect to the STATIC approach and, in more detail, Combined UMD as slightly preferable PLP, for the faded station, among all the proposed techniques.

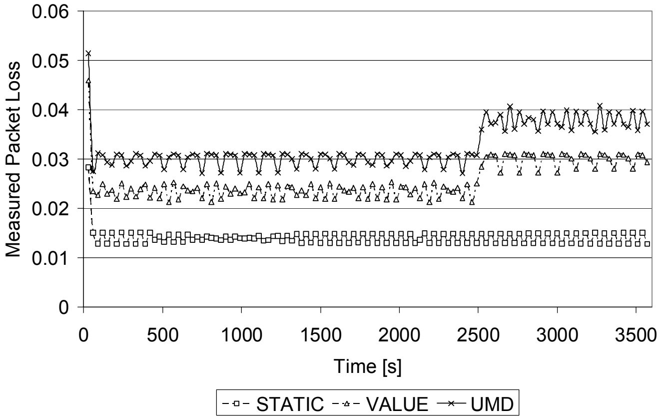

Fig. 3. Packet Loss Performance Comparison (Clear Sky Station).

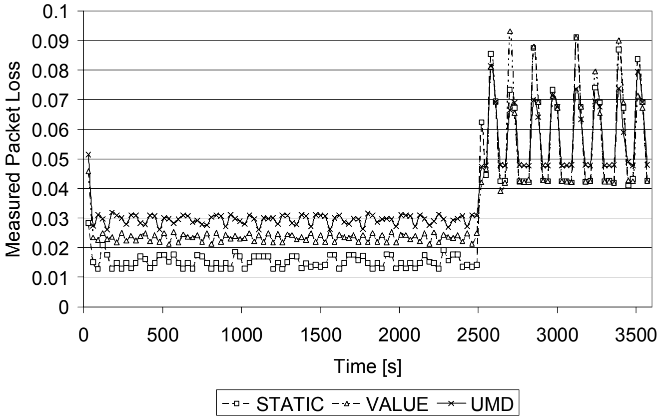

Fig. 4. Packet Loss Performance Comparison (Faded Station).

Figs. 5 and 6 report the same comparison above described but related to the Average Delay. In this case the advantage of the Combined UMD method is really outstanding: in all cases the AD performed by the proposed allocation is better than the other considered approaches and, in particular, considering the faded station (Fig. 6) in the time period $2500 < t \le 3600$, the Combined UMD allows reaching AD of about 200 [ms] while the VALUE has AD of about 300 [ms] and the STATIC approach perform AD of about 700 [ms]. The differences are, obviously, very significant.

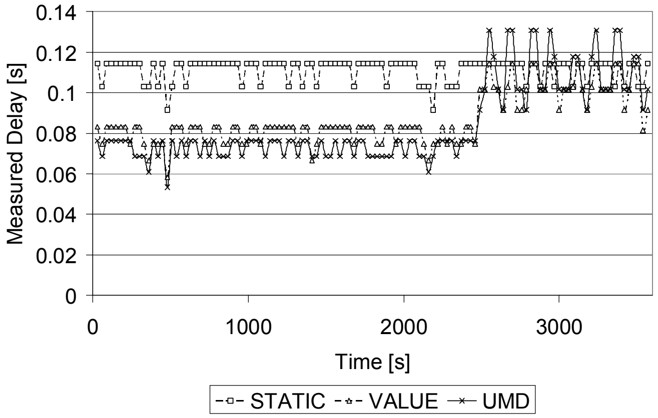

Fig. 5. Delay Performance Comparison (Clear Sky Station).

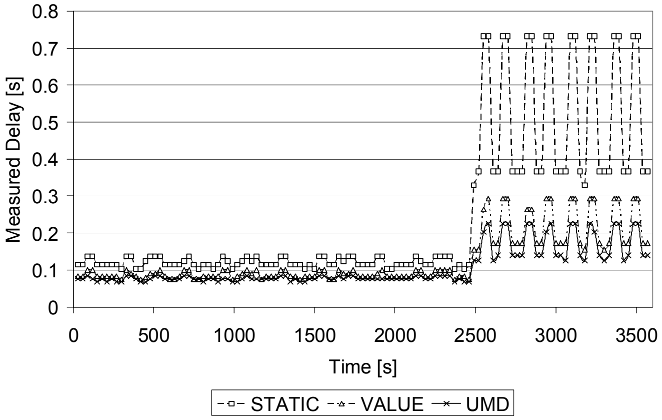

Fig. 6. Delay Performance Comparison (Faded Station).

V. CONCLUSIONS

The paper describes and analyses the Combined UMD allocation scheme for satellite communications. It is suited to be used in Satellite Networks where heterogeneous performance metrics have to be simultaneously optimized. The theoretical framework considered is the Multi – Objective Programming Optimization. The paper investigates the behaviour of the Combined UMD scheme by considering the traffic as a superposition of TCP and UDP sources, opportunely modelled, and compares the results.

REFERENCES

[1] S. Kota, K. Pahlavan, P. A. Leppänen, "Broadband Satellite Communications for Internet Access," Kluwer Accademic Publishers, Boston, 2004.

[2] I. Bisio, M. Marchese, "Minimum Distance Bandwidth Allocation over Space Communications," *IEEE Communications Letters*, vol. 11, no. 1, pp. 19-21, January 2007.

[3] N. Celandroni, F. Davoli, E. Ferro, "Static and Dynamic Resource Allocation in a Multiservice Satellite Network with Fading," *International Journal of Satellite Communications*, vol. 21, no. 4-5, pp. 469-488, July-October 2003.

[4] R. Bolla, F. Davoli, M. Marchese, "Adaptive Bandwidth Allocation Methods in the Satellite Environment," In Proc. *International Communications Conference* (ICC 2001), Helsinki, Finland, June 2001, pp. 3183-3190.