A Dynamic Bandwidth Resource Allocation based on Neural Networks in EuroSkyWay Multimedia Satellite System

P. Camarda(1), M. Castellano(1), G. Piscitelli(1), D. Striccoli(1), G. Tomasicchio(2)

(1) Department of Electrical and Electronic Engineering
Politecnico di Bari, Via Orabona 4, 70126 Bari, Italy
camarda@poliba.it, castellano@deemail.poliba.it

(2) Alenia Spazio, Business Multimedia Unit
Via Bona 85, 00156 Roma, Italy
g.tomasicchio@roma.alespazio.it

Abstract: Advanced traffic management based on the Dynamic Resource Assignment allows a broadband satellite system such as EuroSkyWay (ESW) to dynamically assign the resources of connections. The mechanisms of the dynamic assignment exploit variations of Burstiness exhibited by real time and non real time Variable Bit Rate traffic sources to perform an optimised resource redistribution. The efficiency of the Dynamic Bandwidth Allocation Capability (DBAC) depends on the accuracy of the traffic source description; inaccurate assessment of the arrival process will cause an overhead and a degraded utilisation of system resources.

In this paper a flexible traffic burstiness predictor for dynamic bandwidth resource allocation based on neural network is presented. The approach perform an online estimation of expected resource requests implementing traffic resource assignment by using a sub-symbolic adaptive representation of the traffic source. The achieved results prove that the flexible approach is more effective than the ones based on fixed schemes designed using analytical traffic source description when applied on the satellite terminal ESW system component. An accurate design of neural network architecture and the low-cost industrial availability of this novel technology lead to an optimal trade-off between customer satisfaction in terms of QoS and system resource optimization also in terms of network operator revenue.

1 Introduction

An efficient use of the resources is an important task in communication systems. Resources should be carefully controlled, guaranteeing the needed Quality of Service (QoS) in terms of Cell Loss Rate (CLR), blocking probability, delay, etc. and maximizing, at the same time, the network operator revenues [1].

Satellite systems, even if based on static Bandwidth Allocation, support mechanisms for dynamic solutions to the Resource Management with respect to Real Time Variable Bit Rate and non-real Time Variable Bit Rate traffic sources. Advanced traffic management architectures based on the Dynamic Resource Assignment allow a broadband satellite system to dynamically adapt the resources of provided connections to diversified traffic supported by allowing the resources of each connection to change dynamically.

EuroSkyWay (ESW) is a geostationary satellite network, with regenerative payload based on a circuit-switching scheme [2]. Advanced Resource Management (RM) and Bandwidth Allocation capabilities allow ESW network to fully adapt to the diversified terrestrial traffic by properly allocating the satellite resources for a proper support to efficient switching connectivity services for future broadband data-communication scenarios [3]. ESW adopts a Dynamic Resource Management (DRM) pertaining to dynamic assignment and releases the bandwidth resources required from VBR connections once it has been established. The EuroSkyWay system has been conceived to provide efficient connectivity services to users located within the coverage area of a satellite constellation with high capacity on-board processing payloads. ESW will provide multimedia symmetric and asymmetric, unidirectional or bi-directional, point-to-point or point-to-multipoint connectivity services. The ESW Ground Segment is partitioned in two segments: Control Segment and Traffic Segment. The Control segment includes the Network Operation Centre (NOC) and the Satellite Operation Centre (SOC).
ESW-NOC is composed by the four subsystems: (i) Network Control Centre (NCC) which implements those real-time functions needed for provision of connectivity services, (ii) Network Management Centre (NMC) which implements the overall Network monitoring and control functions, (iii) Customer Care Centre (CCC) which supports the relationships with customers and service management functions, (iv) Data Base Control System (DBCS) which provides a common database facility shared by all NOC subsystems.

As depicted in Fig. 1 the ESW network ground segment architecture uses different Network Elements: Satellite terminals (SaT) for End Users, Provider terminals (PrT) for Service Providers and Gateway terminals (GTW) for Telecom operators interfacing terrestrial networks.

In this paper, a resource requesting scheme based on Neural Network Burstiness Predictor architecture [4, 5] applied in the framework of the EuroSkyWay system is presented. The novel approach, called ARAS (Adaptive Resource Assignment System) is able to compute the expected resource request in real-time, implementing direct adaptive traffic resource management without an analytical representation of the traffic, but using a sub-symbolic description of burstiness. In ARAS, a traffic source is characterized via the number of "active" Frame Units belonging to the ESW’s frame structure; the characterization is based on the prediction. Three MPEG-1 traffic sources have been considered in the experiment. The training data set for the artificial neural network has been collected by means of another MPEG-1 trace. To validate the proposed algorithms, several simulation experiments have been carried out, comparing a fixed scheme approach with the proposed NN-based mechanism.
This paper is organized as follows: Section 2 reports the dynamic resource management architecture implemented in the EuroSkyWay satellite system; in Section 3 the adaptive resource assignment system based on neural network is described; Section 4 and 5 show simulation and numerical results, respectively. Conclusions are given in Section 6.

2 ESW Dynamic Resource Management Architecture

A functional diagram of the ESW Connection Management (CM) and Resource Management (RM) architectures is represented in Fig. 2.

The entities responsible for the RM functionality are placed in the terminals and on-board. The Traffic Resource Manager (TRM) is a portion of the on board baseband processor. On the ESW terminal side a traffic management agent inter-operates with the TRM through an inner Resource Management Entity (RM-E). The Connection Management (CM) functionality is hosted by an NCC sub-system named Connection Control Manager (CCM) which also invokes a Connection Admission Control (CAC) procedure. On the ESW terminal side a Connection Control Agent inter-operates with the CCM through an inner Connection Control Entity (CC-E). Between the CC-E and the RM-E the traffic generated by each source is sustained by means of buffers. The RM functionality is implemented at MAC layer, the CM one is implemented at ESW Layer 3.

The connection acceptance relies on an ESW Connection Profile which states the traffic contract parameters the network will grant to the connection which are: the Service category (A, B, C, D), the Maximum Cell Transfer Delay (Max CTD), the Peak to Peak cell Delay Variation (p-t-p CDV), the Cell Loss Ratio (CLR), the Peak Data Rate (PDR), the Utilisation Factor (UF), the Maximum Burst Size (MBS).
The service categories foreseen from ESW are listed in Tab. 1 with the correspondent relevant traffic features.

<table>
<thead>
<tr>
<th>Service</th>
<th>Connection</th>
<th>Application</th>
<th>UF</th>
<th>Max CTD</th>
<th>p-t-p CDV</th>
<th>CLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Conn. Oriented</td>
<td>CBR</td>
<td>Specified</td>
<td>Highly Sensitive</td>
<td>Highly Sensitive</td>
<td>Slightly Sensitive</td>
</tr>
<tr>
<td>B</td>
<td>Conn. Oriented</td>
<td>rt-VBR</td>
<td>Specified</td>
<td>Slightly Sensitive</td>
<td>Sensitive</td>
<td>Slightly Sensitive</td>
</tr>
<tr>
<td>C</td>
<td>Conn. Oriented</td>
<td>nrt-VBR</td>
<td>Specified</td>
<td>Not Sensitive</td>
<td>Not Sensitive</td>
<td>Sensitive</td>
</tr>
<tr>
<td>D</td>
<td>Conn. Less</td>
<td>ABR/UBR</td>
<td>Not Specified</td>
<td>Not Sensitive</td>
<td>Not Sensitive</td>
<td>Not Sensitive</td>
</tr>
</tbody>
</table>

**Tab. 1 – ESW Service categories**

At connection set-up time a univocal connection identifier is assigned to each connection, the EuroSkyWay Virtual Connection Identifier (EVCI). To support a bi-directional traffic exchange, two different EVCIs must be supplied pertaining to the forward and backward connection (i.e. the connection coming from the calling user to the called one and vice versa).

ESW Dynamic assignment mechanism is suitable to define services provided for traffic source with the following characteristics: rt-VBR, nrt-VBR, no stringent jitter (peak to peak CDV) constraint, high on/off behaviour or Burstiness defined as the ratio: $\frac{T_{ON} + T_{OFF}}{T_{ON}}$ where $T_{ON}$ is the duration of the active time of a traffic source, while $T_{OFF}$ is the duration of the inactive one.

The ESW dynamic assignment mechanism aims at exploiting variations of burstiness exhibited by B and C class sources, to perform an optimised redistribution of those resources to other connections (pre-emption). To this aim the MAC layer foresees a couple of messages which can be transferred in the RM environment between Terminals and the TRM named Capacity Request Messages (CRMs). CRMs can be of two types: Out of Band Capacity Request (OBCR) and In Band Capacity Request (IBCR). Both IBCR and OBCR messages specify the EVCI which they refer to and the resource duration expressed in ESW Frames is 26.5 ms. Although the ESW dynamic assignment mechanism assures, in a simple way, the traffic following in terms of assigned resource, an inaccurate assessment of the $T_{ON}$ will cause an increase in the number of CRMs and a degraded utilisation of unused resources within the system. The number of frames assigned to the traffic source should be strictly related to the appropriate assessment of the activity period that the bursty source will exhibit in the future. A rough solution to assess this value is based on the knowledge for each VBR source, of both the size of the buffers depicted in Fig. 2 and the PDR declared within the relevant ESW connection profile.

---

1. B and C service classes prescribe different constraints about CDV, as shown in Table 1.
2. The ratio has to be intended as a relationship between mean values calculated on a given time window.
3 The Adaptive Resource Assignment System

The architecture of an Adaptive Resource Assignment System (ARAS) is shown in Fig. 5.

![ARAS Architecture Diagram](image)

*Fig. 3 – ARAS architecture*

The Burstiness Predictor module provides the estimation of the resource amount to be required in advance for the system to cope with the high time-varying bursty behaviour that a traffic source may show. Fig. 3 depicts how these modules interact with the other elements of the overall ESW system. The key factors which are addressed by the integration of a Burstiness predictor in the current RM Architecture are the following:

- Improved accuracy in the activity period assessment, due to the fact that the Neural Networks have the ability to learn the behaviour features of a traffic source;
- Reduced CDV, essentially due to a lower usage of the buffers, because the prediction ability reduces the time delay between the capacity request time on the terminal side and the actual availability time of the requested resources after the TRM have granted them.

The predictor design is provided by means of Neural Networks (NNs). NNs are able to discover complex relationships by samples. In terms of modelling the statistical behavior of source traffic to optimize the use of the ESW resources, NN can offer a good representation of the problem without knowing an analytical description of the traffic [6]. The approach here proposed is to employ the Neural Networks as a tool to characterize the *time varying process* of the source and hence to predict its future behavior over a certain monitoring period [7]. One hidden layer feed-forward back propagations neural network is trained to learn the relationship between the characteristics of the source and the future resource amount required to support it. The basic idea is to enable the NNs to predict future samples of this high time-varying process, which is also bursty with significant correlations [8].

A specific description of the above mentioned functional modules will be given in the next sections.
3.1 Extracting Features for Bursty Traffic Source

Three different types of user terminals (hereafter called Satellite Terminals or SaTs) will be supported by the ESW Satellite system. SaTs are classified as A, B or C type depending on the maximum user data rate in the uplink, which is reported in Tab. 2.
Each carrier will be organised according to a Frame structure:

- The Frame Unit (FU), having a length of 680 bits, whose duration depends on the terminal SaT type.
- The Frame, which is the basic unit; it is a set of N Frame Units, with N depending on the terminal SaT type; the Frame has a constant duration of 26.5 ms, regardless of the terminal SaT type.

A single Frame is composed of a number N of FU. The number N and the duration of the FU both depend on the terminal SaT or Carrier Group type data rate, according to Tab. 3:

<table>
<thead>
<tr>
<th>Carrier Group Type</th>
<th># FU per Frame N</th>
<th>FU duration (usec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SaT-A</td>
<td>10</td>
<td>2650</td>
</tr>
<tr>
<td>SaT-B</td>
<td>32</td>
<td>828.125</td>
</tr>
<tr>
<td>SaT-C</td>
<td>128</td>
<td>207.03125</td>
</tr>
</tbody>
</table>

A Frame Unit contains a single ESW cell, having a constant length of 60 bytes (or 480 bits), and a Forward Error Correction (FEC) section added to each ESW cell to increase the immunity against noise.

The Adaptive Resource Assignment approach provides a prediction of the number $N_f$ of Frame Units which will be active in each step $T_S$ with a time span also equal to $T_S$, (see Fig. 5). This time step is $T_s = p \cdot T_{frameESW}$ a multiple of the ESW Frame duration ($T_{frameESW} = 26.5$ms). The time step $T_S$, or Sampling Period, is the time interval between two subsequent predictions. This implies that the maximum frequency for the CRMs’ emissions is now limited by $1/T_S$. Moreover, supposing
\( p \geq 12 \), then the advance time provided by the prediction is about equal to round trip delay between the ESW terminal and the onboard TRM\(^3\).

The Feature Extractor architecture is composed of the following elements:

- a sampler with \textit{Sampling Period} \( T_S \);
- an Active Frame Count Process \( N_f(t, t + T_S) \) which makes an on-line measure of the traffic burstiness through a count of the number of active Frame Units observed during an interval \( T_S \) of a sampled arrival process (see Fig. 6),
- a Shift Register with capacity \( T_m = S \cdot T_S \) storing for each \( T_S \) the above mentioned on-line measure.

\(^3\) This is true if the internal processing delay is neglected.
The architecture shown in Fig. 6 refers to each source. In the sampling window $T_S$, a multiple of the frame duration, the Sampler tags each Frame Unit with a binary digit in accordance with the active or inactive state. As it will be further explained, $T_S$ is the time in advance during which the predictor provides a number of Frame Units assessed between 0 and $M \cdot p$ (where $M$ is the number of Frame Units in an ESW Frame; i.e., $M = 128$ for ESW Sat-C terminal). An Active Frames Count Process counts how many Frame Units have been active in $T_S$. These counts are loaded in a Shift Register spanning a time duration of $T_m = S \cdot T_S$. The shifting mechanism at the Shift Register level implements a sliding window large $T_m$ moving with step of $T_S$. The NN provides, from the past bursty behavior of the source $S_i$, the expected number of Frame Units which will be active in the next $T_S$ time period.

The larger accuracy that a neural predictor exhibits in the frame assessment with respect to rough size buffer mechanisms guarantees a better utilization of the resources among all traffic sources.
3.2 The Neural Network module

The Neural Network (NN) used in the ARAS architecture calculates a bursty correlation of the active Frame Units diffused in the arrival process on the time interval $T_m = S \cdot T_s$. The NN learns the relationship between the characteristics of the traffic source arrival process and the related estimation in advance of the number of active Frame Units $(T_{ON})_{N_f}$, with $0 \leq N_f \leq M \cdot p$ required to support it in the next interval $T_S$. The adopted NN model is shown in Fig. 7 and is assumed to be fed from the previously described feature vector to implement the detector block.

![Neural Network topology](image)

**Fig. 7 - Neural Network topology**

A three-layered feed-forward Neural Network in which units are arranged in fully connected layers and connections are allowed only between contiguous layers is taken into account. Each neuron $h,(h=1,...,N_h)$ belonging to the hidden layer receives as input the output $X_l,(l=1,...,N_i)$ of all the neurons $S$ of the input layer, which it is connected to by $w_{hi}^{(1)}$ synaptic strengths. On the other hand, the neuron $h$ is also connected, with strengths $w_{hi}^{(2)}$ to all the neurons $i,(i=1,...,N_O)$ of the output layer $O$. The transfer function of the neurons is described by the sigmoid function:

$$g(x, \theta) = \frac{1}{1 + \exp(-\beta(x - \theta))}$$

where $\theta$ and $\beta$ are the neuron threshold and the gain factor, respectively. The synaptic matrix $W$ is trained by using the back-propagation techniques, in the incremental learning mode: each weight $w$ is updated after every pattern $X$ has been forwarded to the NN inputs, as follows:
\[
\Delta w = - \eta \frac{\partial E[W]}{\partial w} + \alpha \Delta w^{(old)}
\]

where \( E[W] \) is the distance between net output \( Y(X) \) and the known expected answers \( T(X) \). The momentum term \( \alpha \Delta w^{(old)} \) is introduced in such a way that each connection value \( w \) tends to change in the average downhill direction, avoiding sudden oscillations. The evaluation system has been designed by using a multi-layered NN with one hidden layer and \( N_h = 2N_f + 1 \) hidden neurons, according to Kolmogorov’s general theorem [9]. The output layer consists of \( N_O = 1 \) elementary processing unit. The network parameters used for a suitable implementation of the gradient descent algorithm are: learning rate \( \eta = 1 \), momentum \( \alpha = 0.9 \) and gain factor \( \beta = 1 \) [10].

The neural network produces a prediction system which, starting from the analysis of the feature vector in the appropriate feature space, makes the final decision about the event providing the indication of the amount of resource request which will be dynamically assigned.

Fig. 8 illustrates the basic idea in training the NN to act as a predictor. For each input, the network produces an output which is then compared to the desired output (target) [11]. The arrival process is represented by the data \( [N_f(i+1)] \) which is the NN target output. It represents the number of active Frame Units to the time \( i+1 \), measured in \( T_S \) intervals, a multiple of the frame ESW (\( T_S = p \cdot T_{frame ESW} \)). The data \( [N_f(i+1)] \) provides the NN with the arrival process information from which the predictions will be made. The NN predicts the arrival process variations by exploiting the inherent correlations that exist.

---

**Fig. 8 – The neural network training phase system architecture**
For training purposes, the NN input data vector \([N_f(i-S+1), N_f(i-S+2), \ldots, N_f(i-1), N_f(i)]\) contains \(S\) elements representing the measured number of active Frame Units in the preceding \(S\) intervals \(T_S\). The duration of the measurement period \(T_m = S \cdot T_S\) must be chosen such that the effect of correlation is well captured. The NN then tries to match the target output data \([N_f(i+1)]\) with its predicted output data \([N_f(i+1)]\) by adjusting its weights. It then follows that when the input data bypass the delay unit, the output \([N_f(i+1)]\) is a prediction of the number of active Frame Units in the future \(T_S\) step. The delay unit, shown in Fig. 8, delays its input \([N_f(i+1)]\) of one \(T_S\) time step. Assuming that the NN requires a negligible amount of time to compute output, the NN, after training, provides estimations for the value of active Frame Units \([N_f(i+1)]\) one step in advance. For each epoch, test data set is fed in input to the network and the Root Mean Square error (RMS) is evaluated. The training phase is stopped when the overtraining effect [10] is observed and the best synaptic matrix is frozen.

### 3.3 The Decision Logic module

The Decision Logic module implements a decision making facility able to trigger the proper CRM at the RM-E of the terminal. Both the CRM formats require the relevant EVCI and the predicted \(N_f\) to be specified. Each decision logic module acts on the basis of the following information:

- traffic contract parameters within the relevant ESW Connection Profile
- predicted number of active Frame Units \(N_f\) in the next period \(T_S\)
- Traffic source Buffer load status

The output of this module involves the invocation or not of a primitive such as \(\text{Make\_Resource\_Req(EVCI}_i, N_f)\) to be processed at MAC layer from the RM-E entity within the terminal.

### 4 Simulation Results

Extensive simulations were performed to obtain the NN’s data set, for training phase and also to assess the performance of various NN architectures. The Frame Unit arrival process, resulting from the packetized video source, was simulated. Assuming a video source packetized for using ESW network, in the case of Sat-C terminal type, the maximum number of active Frame Units is 128 in each ESW frame; an ATM cell, or Frame Units ESW, has size 53 bytes with 48 bytes of payload, and the frame ESW interarrival time \(T_{frame \_ESW}\) is 26.5ms. An important parameter which must be
defined is the sampling period $T_S$. The choice of this parameter is influenced by the type of prediction of the traffic and also by the system architecture. In this application it has been found to be around 400ms. NN is trained to predict the number of active Frame Units for the next sampling period $T_S$ based upon the $S$ past values of the number of active Frame Units count process.

4.1 Traffic Source

The source data utilized for our analysis consist of portions of video streams codified with the MPEG-1 video standard compression [12]. The particular type of compression adopted generates a Variable Bit Rate (VBR) type of traffic, in which there is a high variability between adjacent frames of the video stream. In particular, each of the video frames is transmitted in a constant time interval, depending on the transmission rate of the video frames (the transmission rate adopted in this case is 25 frames per second; in this way, each video frame is transmitted in 40 milliseconds), while in each video frame a different number of bit is present. This is due to the particular nature of the standard compression adopted, which generates three types of video frames: I-frame, P-frame and B-frame [13]. The I-frames (Intra-coded frames) are codified in an independent way and they do not depend on other pictures in the video sequence. The P-frames (Predictive-coded frames) are coded exploiting dependencies from the temporally preceding I-frames or P-frames, while B-frames (Bidirectionally predictive-coded frames) exploit correlation with preceding and upcoming I-frames or P-frames. It is obvious that the I-frames have the higher number of bit, if compared with P-frames and B-frames, because each particular of the corresponding picture is coded independently from other pictures. Similarly, the P-frames present a number of bits lower than I-frames, but higher than B-frames, because they are coded exploiting dependencies from temporally preceding I-frames or P-frames only. B-frames have a little number of bit because they exploit dependencies with preceding and upcoming I-frames or P-frames. In any case, the number of bits present in each frame of the video sequence can vary, based on the dynamic evolution of the scenes. In Fig. 9 there is an example of a portion of 32000 video frames, extracted from a talk show, compressed utilizing the MPEG-1 algorithm. The high variability of the video frame’s size can be noted.

14
I-frames, P-frames and B-frames are transmitted in a well defined pattern that repeats in a periodic way until the end of the entire video stream. A typical frame pattern adopted in the case of MPEG-1 standard is the following: \texttt{IBBPBPBPBPBPBBP...} and is reported in Fig. 10:

![Frame pattern of a Group Of Pictures](image)

Each of these groups of video frames is called Group Of Pictures (GOP). In this case, the length of each GOP is 12 video frames.

It is known [14, 15] that the MPEG video streams exhibit properties of self-similarity and Long Range Dependence (LRD). The self-similarity phenomena can be intuitively explained observing that MPEG video streams exploit structural similarities and correlations on different time scales, ranging from few milliseconds to hours of video streams. In particular, the bit rates of a video stream present a high amount of correlation, also for relatively large time lags, and a slow decaying
correlation curve. To take into account the self-similar properties and the slow decaying correlation of the video streams, the Hurst parameter $H$ has been introduced. It can be estimated with different methods, in both cases of finite or infinite variance of analyzed data, and can assume all the values between 0.5 and 1. In this work, we will utilize for our experiments three different types of video streams: James Bond, Jurassic Park and The Simpson's. For each of them, the Hurst parameter, obtained utilizing the R/S method [16], is reported in Tab. 4. The values of $H$ will give us an idea of the correlation between the video frames in a video stream.

<table>
<thead>
<tr>
<th>SEQUENCES</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAMES BOND</td>
<td>0.90</td>
</tr>
<tr>
<td>JURASSIC PARK</td>
<td>0.88</td>
</tr>
<tr>
<td>THE SIMPSON’S</td>
<td>0.96</td>
</tr>
</tbody>
</table>

*Tab. 4 – Hurst parameter for the three types of films*

The first step is to establish the number of active Frame Units in a temporal window generically chosen of $M$ frames ESW. In each of these windows, we have to count the total number of Frame Units that contain some video data (these Frame Units will be marked as 1 to indicate the fact that there is some data in them, while if a frame unit is empty, it will be marked as 0). For each window, the total number of active Frame Units is counted and then stored in a text file. All the portions of video streams utilized in our experiments have a length of 40.000 video frames (3.333 GOPs).

4.2 Performance Metrics

In the performance metrics we introduce the resource utilization factors which will be used in the next sections. For the fixed scheme based on a buffer with threshold $\theta$ we define the utilization factor $F_B$. This scheme is based on two bandwidth levels: if the buffer is filled under its threshold $(used \_capacity)_\theta$ then the assigned bandwidth corresponds to the buffer threshold $L_\theta$; if the buffer is filled over its threshold $(used \_capacity)_B$, the assigned bandwidth is equal to the full buffer capacity $L_B$. 

16
The $F_B$ factor, expressed in percentage, is:

$$F_B = \frac{1}{T_{TOT}} \sum_{t=0}^{T_{TOT}} \left[ \frac{(used\_capacity)_B}{L_\theta} + \frac{(used\_capacity)_\theta}{L_B} \right] \cdot 100\% \quad (4.1)$$

where $T_{TOT}$ is the total duration of the video stream.

The $F_{NN}$ is the utilization factor for the flexible scheme based on NN. Let $\theta_t$ and $(used\_capacity)_t$ respectively the bandwidth predicted and bandwidth utilized at the $t^{th}$ time step.

It will be:

$$F_{NN} = \frac{1}{T_{TOT}} \sum_{t=0}^{T_{TOT}} \left[ \frac{(used\_capacity)_t}{\theta_t} \right] \cdot 100\% \quad (4.2)$$

It can be noted that the upper bound of the used capacity $(used\_capacity)_t$ is equal to $\theta_t$.

Let now $(needed\_capacity)_t$ the bandwidth required at the $t^{th}$ time step. Let us define:

$$d_t = \theta_t - (needed\_capacity)_t \quad (4.3)$$

If $d_t \geq 0$ then the bandwidth is overestimated and a waste of bandwidth will occur. If $d_t < 0$ the bandwidth is underestimated and there will be data losses. The total amount of Lost Frame Units (LFU) is taken into account in the following loss factor:

$$LFU = \frac{total\ active\ Frame\ Units\ discarded}{total\ active\ Frame\ Units} \cdot 100\% \quad (4.4)$$

Moreover, a study on the NN behaviour can be done adding a fixed amount of bandwidth $\sigma$ to the NN predicted bandwidth\(^4\) at the aim to reduce the total amount of lost Frame Units. A merit figure can be obtained plotting the $F_{NN}$ and $LFU$ factors versus $\sigma$, where $\sigma$ is expressed in terms of percentage of the satellite terminal capacity.

4.3 Neural Network Training Phase

The training data set for the NN is now collected. This data set comes from a variable bit rate video trace codified using MPEG-1 algorithm. The movie contains a mixture of material ranging from low complexity scenes to scenes with high action.

\(^4\) The maximum associable bandwidth for the single step is always fixed within 2048 Frame Units.
The data set has been chosen imposing the value of $p=16$ in order it covers at least the Round Trip Time. Therefore the period between one prediction and the next has been chosen equal to $T_s = p \cdot T_{frameESW} = 16 \cdot 26.5\,ms = 424\,ms$. On the other hand, the duration of the measure period $T_m$ has been chosen imposing $S=4$, which settles down the dimension of the sliding window implemented with the Shift Register. The $S=4$ value has been chosen after a series of experiments carried out on different NN architectures, as shown in Fig. 11 and Fig. 12. In Fig. 11 it can be noted that the training error observed for a large number of epochs is almost the same for NN architectures with the input neurons greater than 3. In this case the less complex NN architecture can be chosen.

![Fig. 11 – Rate of convergence in the training phase](image-url)
Fig. 12 - Rate of convergence for the selected NN(4-9-1) test phase: the overtraining effect is noted at the 2000th epoch.

In the sequel the selected NN(4-9-1) will be used with the synaptic matrix frozen at the 2000th epoch.

5 Experimental Results

In this section the effectiveness of the proposed ARAS is studied by demonstrating its capability to produce the prediction of the $N_f$ number of active Frame Units in the next sampling period $T_S$. To this purpose the following two experiments are performed:

1. NN proposed is validated as a tool for burstiness prediction using different real video traces codified with MPEG-1 algorithm;
2. Neural network architecture ARAS is compared with the buffer with threshold mechanism.

Experiment 1: in this experiment we intend to show how the neural network, after the training phase based on an MPEG-1 video trace, generalizes its learning. Four MPEG-1 video traces are examined. Fig. 13 shows the typical behaviour of the NN when compared with a real video trace.
In Fig. 13 it is noted that the estimated bandwidth follows well the real one. In other words, the neural network can predict the number of the active Frame Units present in the next time step $T_s$. The ESW bandwidth resource management will be able to use the estimated resources, by IBCR/OBCR signalling, with an advance of $p$ ESW frames, taking into account the Round Trip Time delay.

With the purpose to evaluate the effectiveness of the predictive approach, in terms of reliability, we introduce the step by step comparison between the needed number of the active Frame Units $N_{f,\text{needed}}$, requested by the video trace, and the predicted number of the active Frame Units $N_{f,\text{predicted}}$, estimated by the neural network. In such a way, we can extract a reliability index $\alpha$, given by the step by step relationship:

$$\alpha = \frac{N_{f,\text{needed}}}{N_{f,\text{predicted}}}$$  \hspace{1cm} (4.5)

The behavior of the reliability index $\alpha$, the distribution of the $d_i$ indexes as defined in (4.3) and the $F_{NN}$ and $LFU$ factors are shown in Fig. 14 for the “James Bond” video trace, in Fig. 15 and Fig. 16 for “Jurassic Park” and “The Simpson’s” video traces, respectively.
Fig. 14 – “James Bond” video trace: a) the Active Frame Units profile, b) the Reliability index and c) the distribution.

Fig. 15 – “Jurassic Park” video trace: a) the Active Frame Units profile, b) the Reliability index and c) the distribution.

Fig. 16 – “The Simpson’s” video trace: a) the Active Frame Units profile, b) the Reliability index and c) the distribution.
In Fig. 17 are shown the indexes performed by the Neural Network using the same video trace “Talk show” on which the selected NN has been trained. The NN validation emerges comparing Fig. 17 with Fig. 14, Fig. 15 and Fig. 16 since there is a good agreement among them.

\[ \text{Fig. 17 - Temporal distribution, Reliability index and Distribution for training’s video trace, the “Talk Show”} \]

It is pointed out that if \( \sigma \) increases, a relevant decrease of the \( F_{NN} \) and \( LFU \) factors arises.

**Experiment 2:** in this experiment we intend to show how the ARAS architecture, applied to the ESW Sat-C terminal, provides a higher resource utilization factor than a static mechanism, like the buffer with threshold. The buffer system is implemented to dynamically change the capacity of the single connection without repeating the set-up procedure. In this case, we also assume that the bandwidth assigned in each time interval of an ESW frame is instantaneously exploited by that frame, without any kind of delay between the bandwidth negotiation and the bandwidth assignment. Given this assumption, the factor \( F_B \) has to be considered an upper bound that can be reached only disregarding the Round Trip Time delay. The dynamic change of bandwidth requirements is useful for an optimal resource allocation while respecting the guarantee of correct delivery of the video stream. In particular, it has been supposed that a single source, composed by 40000 video frames of a VBR video stream compressed with MPEG-1 algorithm, enters a buffer with a fixed threshold. The buffer capacity is assumed to be of a single ESW frame so that, during the video transmission, the buffer is always empty when a new ESW frame arrives. In this study the satellite terminal capacity corresponds to the maximum bandwidth that can be assigned to the video source, but it can vary dynamically during the video transmission. The bandwidth, the terminal capacity and the buffer threshold in this case are expressed as number of Frame Units in an ESW frame duration.
The terminal capacity is assumed SaT-C and is 128 Frame Units. The SaT-A and SaT-B capacities have not been taken into account in this study since the losses of video frames are too high. Given the VBR nature of the video source, not all of the Frame Units in each ESW frame will be filled with data.

In Fig. 18 the curves expressing the values of $F_B$ varying with the position of the buffer threshold for different types of video traces are shown. For each video trace an input data of 40000 frames, equal to approximately 30 minutes of film, have been considered with a transmission rate of 25 video frames per second. The buffer threshold has been varied from a minimum of 1 Frame Unit to a maximum of 128 Frame Units. For each of the threshold values, the corresponding value of $F_B$ has been computed.

![Fig. 18 – $F_B$ factor for the “James Bond”, “Jurassic Park” and “The Simpson’s” films, on SaT-C](image)

From the analysis of the previous figure, it can be noted that $F_B$ reaches a maximum in correspondence to a given buffer threshold value. The maximum value of $F_B$ is not the same but varies with the particular considered video trace.

The same video traces then feed the neural network architecture. In Tab. 5 a comparison is reported between the utilization factors obtained with the two different approaches. The buffer’s utilization factor is derived in correspondence to the optimal threshold while the NN’s utilization factor is associated with the LFU loss factor.
## 6 Conclusions

In this paper a flexible Adaptive Resource Assignment System for the EuroSkyWay broadband satellite system has been presented in the complex scenario of the transmission of bursty VBR video sources. This approach when applied to the ESW satellite terminal component is able to compute in advance the needed bandwidth resources in an efficient way. The results prove that the flexible approach is more effective than the ones based on fixed schemes designed using analytical traffic source description.

An accurate design of neural network architecture and the low-cost industrial availability of this novel technology lead to an optimal trade-off between customer satisfaction and system resource optimization also in terms of network operator revenue.

### References


