Human-centric Computing and Information Sciences

CrossMark

Video scene identification and classification for user-tailored QoE in GEO satellites

Polychronis Koutsakis[1]*, Irene Spanou[2] and Aggelos Lazaris[3]

*Correspondence:
p.koutsakis@murdoch.edu.au
[1] School of Engineering
and Information Technology,
Murdoch University, Science
and Computing Building
245, SC1.012, 90 South Street,
Murdoch, WA 6150, Australia
Full list of author information
is available at the end of the
article

Abstract

Satellite networks offer an efficient alternative where no terrestrial networks are available, and can offer a cost effective means of transferring data. Numerous proposals have addressed the problem of resource allocation in geostationary (GEO) satellites, and have been evaluated via various performance metrics related to user Quality of Service (QoS). However, ensuring that on average the users' QoS requirements are satisfied does not guarantee the satisfaction of their individual Quality of Experience (QoE) requirements, especially in the case of bursty video users. This paper proposes the use of video scene identification and classification for traffic modeling at the scene level in order to improve resource allocation and user QoE in GEO satellite networks. The proposed call admission control and medium access control framework makes decisions based on available bandwidth, long-term and short-term user satisfaction and the revenue that the provider may gain.

Background

The use of satellite links ensures connectivity where no terrestrial networks are available and can offer an efficient and cost effective means of transferring data. Satellite systems offer wide coverage, flexibility and reliability, as they rely on permanent terrestrial infrastructures just for the gateway [1].

Although GEO satellites have been a subject of substantial focus by the telecommunications industry, they still pose significant network design problems, such as long propagation delays, the limited bandwidth to be shared among many users, and limitation in power. Hence, efficient scheduling at the medium access control (MAC) protocol level is needed to properly allocate network resources, especially when coping with bursty multimedia users. The MAC protocol needs to be combined with an equally well-designed call admission control (CAC) scheme, which will not only serve the traditional role of CAC mechanisms (i.e., to prevent traffic overload) but also to maximize the satellite provider's profit without jeopardizing the Quality of Service (QoS) offered to multimedia users and leading to user dissatisfaction.

The design of MAC and CAC mechanisms for GEO satellites that will be able to handle bursty video traffic is becoming even more important as GEO satellites are expected to play a major role in 5G networks. More specifically, satellites will be used to off-load

Springer Open

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 2 of 18

traffic from the terrestrial networks, especially video traffic which is the largest contributor to the spectrum demands. This can be achieved by traffic classification and intelligent routing to reduce the demands on the terrestrial spectrum [2]. Still, as explained in [3], due to the problem of large propagation delays, a significantly larger focus has been given to the scheduling problem, at the MAC layer, in GEO satellite networks, than to call admission control. Most of the existing work in the field has not considered bursty video traffic, which calls for a very efficient CAC in order to guarantee high network throughput and the user Quality of Service and Quality of Experience (QoE).

Towards this goal, we have proposed FPRRA [3], a MAC and CAC framework for a digital video broadcasting return channel satellite (DVB-RCS) system, which makes decisions after taking into account the provider revenue. The work in [3] considered an on-board processing (OBP) system [4], where the Network Control Center (NCC) is located onboard the satellite so that it takes the requesting earth station only one round-trip time (RTT) plus the processor/queuing processing time to receive the reply for its reservation request. The reason that in most current broadband satellite access systems the scheduler is on the ground has to do with the computational requirements being less restrictive [5], but in [3] it was shown that the proposed framework added low computational complexity, therefore the "ideal" choice of having the scheduler on-board can be implemented.

The work in [3] used the Discrete Autoregressive Model of order one that was shown in [6] to be highly accurate. However, this accurate prediction is not possible for all types of video sequences, and even when it is, it often involves a higher degree of complexity which would incur additional computational requirements for an OBP system. For this reason, in [7] we studied the efficiency of our scheme in the absence of accurate multimedia traffic prediction, by implementing our DAR(1) modeling approach on video (not videoconference) traces, and studying the effect that the lack of high modeling accuracy had on our framework, in terms of bandwidth utilization and provider revenue. The notion of user satisfaction was also introduced in [7]. It was formulated as user irritation, similarly to the work in [8] for cellular networks. Two different definitions than the ones in [8] were proposed in [7] for the short-term user irritation factor (SUIF) in GEO satellite networks. The respective results were compared and their relation to the proposed resource pricing approach was discussed. The results in [7] showed that, in spite of the much lower traffic modeling accuracy for video sources in comparison to videoconference sources, the proposed scheme was still able to outperform other schemes from the literature.

This paper is organized as follows. In "Related work", related work on video traffic modeling and on modeling oriented to scene detection in particular is presented. The section also briefly presents recent work of ours [9] on a new hybrid video traffic model for MPEG-4 video traces, which significantly improves for video sources the accuracy of the DAR(1) model originally used in FPRRA. A vast number of traces has been used and modeled in [9], most of them different and burstier than the ones used in [7], therefore leading to more strenuous traffic scenarios for the system to handle. For the traces studied, the DAR(1) model does not provide satisfactory results, and hence a new modeling approach of low complexity is needed.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 3 of 18

"The proposed scheme" presents and discusses the main contribution of this work, which is the introduction of the notion of long-term user satisfaction (Irritation) into the decisions of the call admission control scheme of FPRRA. The basic idea of the proposed scheme is that the decision of whether to accept or not a new user in the network must rely on a combination of factors, including the available bandwidth, the network users' long-term Quality of Experience and the revenue that the provider may gain. We take advantage of the new model's ability to distinguish between high and low activity scenes, in order to change the definition of the Sigmoid function used in [7, 8] for representing user satisfaction.

In "Simulation results and discussion" the results of the proposed scheme are compared against a number of schemes from the literature ([10–12]) and against the work in [7], both in terms of results and conceptually. The new scheme is shown to provide significantly better QoS and QoE to video users over GEO satellite networks. "Conclusions" presents our conclusions and goals for future work.

Related work

The problem of video traffic modeling has been extensively studied in the literature, with tens of models focusing on each video-encoding standard. The first autoregressive (AR) model was introduced in [13] and it was followed by other variations such as [14, 15]. A Discrete Autoregressive Model for MPEG-4 videoconference traffic was presented in [6]. In [16] the authors classify video shots in different classes based on their texture and motion complexity, and then each shot class is described with a different autoregressive model. In [17] an MRP transform-expand-sample (TES) is evaluated, which is characterized by higher computational complexity. In [18] a finite-state Markov chain is used to model one- and two-layer scalable video traffic based on the assumption that I frames follow a Gaussian distribution, and an AR model of order 1 is used to model the P-frames. In [19], a Gamma–beta-auto-regressive (GBAR) model is introduced for H.261 videoconference traffic, where each video sequence has Gamma marginal distribution and geometric autocorrelation. In [20], a variation of the GBAR model is proposed that takes into account the group-of-pictures (GoP) cyclicity. In [21], wavelet modeling is used to model the distribution of I frames and a time-domain model for P/B frame sizes, for MPEG-4 and H.264 videos. In [22], a neural network approach is utilized for modeling the I, P, and B frames of MPEG-1 and MPEG-4 video separately.

The problem of video scene detection has also attracted significant attention from the research community for two reasons: (a) it can be used for pure video segmentation [23, 24] which has applications in storing, processing or analyzing the semantics of videos, and (b) it can be used in video traffic modeling in order to improve the model's efficiency [25, 26]. The interested reader can refer to [27] for a more complete treatment of the scene detection literature.

In [9] we proposed a scene change detection-based Discrete Autoregressive Model, which was tested on more than sixty different long sequences of MPEG-4 encoded videos, from a publicly available library [28]. Indicative results for some of the 10 traces which we will proceed to use in our satellite CAC/MAC framework study are presented below. The first three columns of Table 1 present the statistics for each trace. The acronyms {HQ, LQ} stand for high quality and low quality, respectively.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15
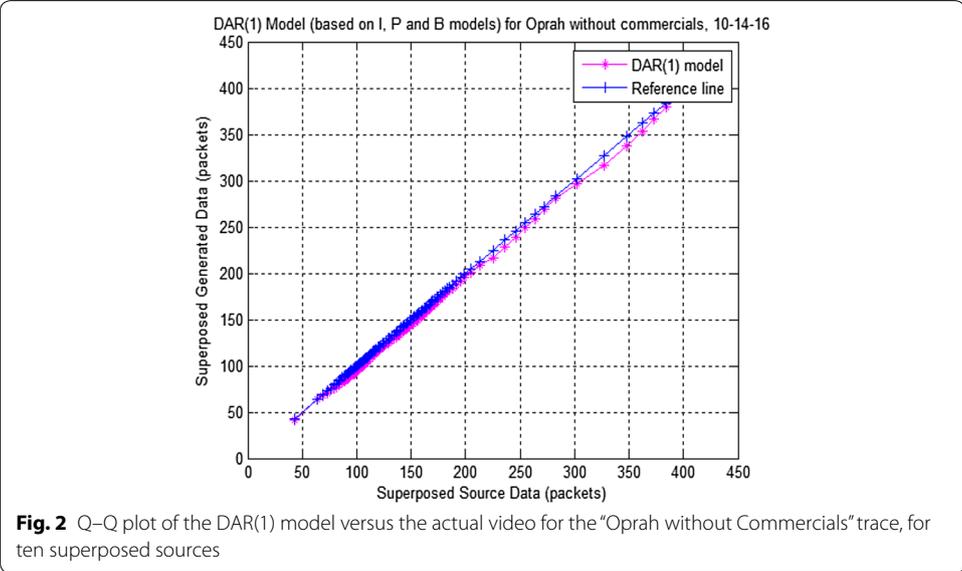
Page 4 of 18

Table 1 Trace statistics
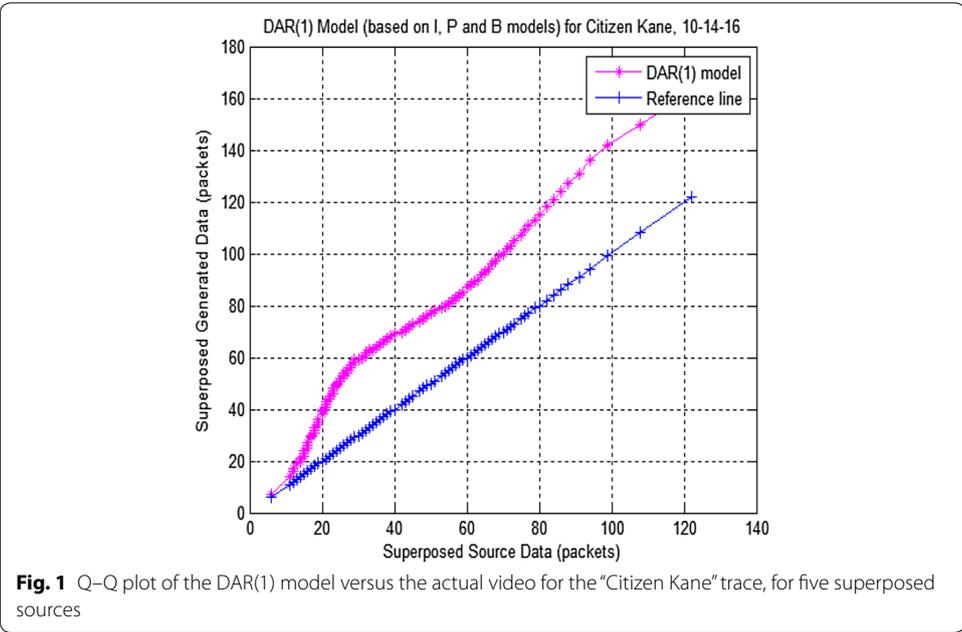
Movie	Mean bit rate (Mbps)	Peak bit rate (Mbps)	Standard deviation (Mbps)	Initial revenue weights and (q) value
Die Hard One (HQ)	0.665	3.109	0.287	2.73 (q = 5%)
Citizen Kane (HQ)	0.457	3.992	0.321	2.52 (q = 10%)
Aladdin (LQ)	0.210	2.243	0.213	1.83 (q = 50%)
Aladdin (HQ)	1.316	5.414	0.441	3.15 (q = 1%)
Tokyo Olympics (HQ)	0.399	3.387	0.317	2.18 (q = 25%)
Tonight Show with Commercials (LQ)	0.267	2.467	0.428	1.96 (q = 40%)
Oprah without Commercials (LQ)	0.183	1.826	0.201	1.47 (q = 80%)
Die Hard One (LQ)	0.122	1.655	0.135	1.65 (q = 65%)
Silence of the Lambs (LQ)	0.080	2.028	0.111	1.23 (q = 95%)
O Brother DVD (LQ)	0.076	1.250	0.110	1.00

For high quality (HQ) traces, the quantization is 04-04-04, while for low quality (LQ) traces, the quantization is 10-14-16

Initially, we attempted to model the traces with the use of the DAR(1) model, similarly to [3, 7]. DAR(1) provides an easy and practical model based only on four physically meaningful parameters (mean, peak, variance and the lag-1 autocorrelation coefficient), therefore it uses parameters which are either known at call set-up time or can be measured without introducing much complexity in the network. We built a separate model for each video frame type (I, P, B); then, these models were used to generate I, P and B frame sizes according to the GoP pattern of each original trace, hence creating a synthetic model for a single trace, and a number of the generated traces were superposed in order to assess the modeling results. We used Q–Q plots [29] for assessing the results, and plotted the 0.01-, 0.02-, 0.03-,… quantiles of the actual video frames' sizes versus the respective quantiles of the respective DAR(1) model, for a superposition of traces. Indicative results are presented in Figs. 1, 2. The modeling accuracy varied from bad (Fig. 1) to good (Fig. 2), with the vast majority of the studied cases leading to mediocre results, i.e., the points of the Q–Q plot failed to fall along the 45° reference line.

Because of the largely unsatisfactory results of the DAR(1) model, we were led to the creation of a hybrid model combining the DAR(1) model with a scene-based Markov chain model, similar to the one presented in [25] (which was chosen due to its simplicity, since it operates at the frame-size level only, similarly to the DAR(1) model, without analyzing the blocks, color histograms, or other information of the video frames). The main concept of that work is the idea of dividing each video trace into scenes, and then classifying the detected scenes into low or high-activity ones. For each movie used in this study, a 2-state (high, low scene activity) Markov chain model was implemented in order to determine the number of low and high scenes of the modeled source. For every scene in both categories, the number of I, P and B frames in it was determined and finally the DAR(1) model was ran, for every case. Details on the steps followed can be found in [9].
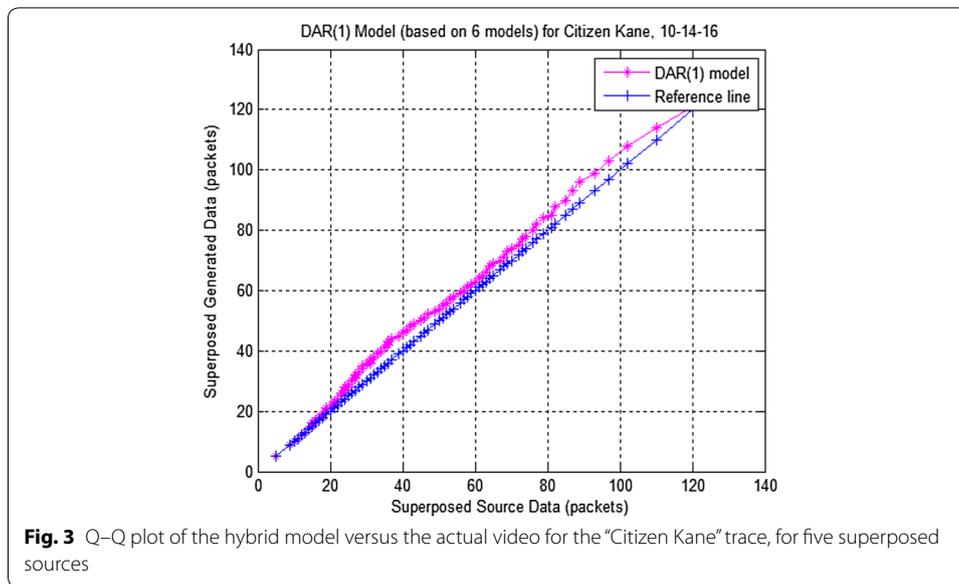
The next step, after scene identification, is scene classification. If a scene has an average bit rate that is greater than the average bit rate of the whole movie, then it is classified as a high-activity scene, else it is classified as a low-activity scene. The average bit rate of every scene was calculated as the number of bits transmitted during the scene divided by the scene duration. Each low and high activity trace was divided into their respective I, P and B frames; hence, every movie was "split" into six subtraces: the low- and

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 5 of 18

Fig. 1 Q–Q plot of the DAR(1) model versus the actual video for the "Citizen Kane" trace, for five superposed sources

Fig. 2 Q–Q plot of the DAR(1) model versus the actual video for the "Oprah without Commercials" trace, for ten superposed sources

high-activity I frames, the low- and high-activity P frames and the low- and high-activity B frames.

After calculating the mean and variance of the frame sizes of each of the $10 \times 6 = 60$ subtraces under study and determining all the distribution parameters, we ran Kullback–Leibler [30] and Kolmogorov–Smirnov [29] tests and generated Q–Q plots, in order to determine the best fit for all cases and use it to separately model each subtrace via DAR(1). Then, the separate models were used to create new synthetic traces. Again, more details can be found in [9].

An indicative result of the hybrid model is shown in Fig. 3, where it is clear that the modeling accuracy improves very significantly with our approach, when compared with Fig. 1.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 6 of 18

Fig. 3 Q–Q plot of the hybrid model versus the actual video for the "Citizen Kane" trace, for five superposed sources

Based on these results, we proceeded to use the model in order to precompute various traffic scenarios. The precomputation, along with the online simulation, is made based on the traffic parameters declared by the video sources at call setup. These parameters are used for the "identification" of the source as a user adopting a specific "mode", i.e., a set of traffic parameters. Hence, the sets of traffic parameters presented in Table 1 are denoted as "modes" for the satellite video users, therefore ten "modes" are used. Each "mode" represents a specific Quality of Service that a user wishes to get. Users choose one of the ten "modes" with equal probability (10%). Also, the "modes" are divided in three groups (of 3, 4 and 3 "modes", respectively) as shown in Table 1. In each group the order of the "modes" represents their quality within the group, which is a result of their trace statistics. The placing of each "mode" in a specific group is necessary for our CAC mechanism, discussed in "CAC based on long term user irritation".

In the context of CAC, it should be noted that the decision of admitting or rejecting a new call in the network should be made by the provider not only based on the capacity needed to accommodate the call, but also on both: (a) the satisfaction/irritation that the admission/rejection of the new call will cause to the user, and (b) the revenue that the admission of the new call will provide. For this reason, the idea of dedicating bandwidth to higher-paying users who may often not need it in its entirety, can easily lead the network to bandwidth starvation and cause irritation to lower-paying users, hence leading to customer attrition.

In [8], the authors presented a two-level resource management scheme for cellular networks, which had both a CAC and a MAC component. Both components used the user irritation factor (UIF) in order to make decisions on the resource allocation in the network, however pricing was not considered in that work. Two irritation factors, the short term user irritation factor (SUIF) and the long term user irritation factor (LUIF), were defined. SUIF measures the delay that the user is ready to suffer prior to which the user decides to cancel a particular request, and LUIF determines the grade of irritation of the user resulting from repeated violation of SUIF thresholds. The work in [8] uses a

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 7 of 18

Sigmoid function which it correlates with the SUIF and LUIF metrics. For a random variable x representing a service parameter like delay, the corresponding satisfaction, $U(x)$ decreases with increasing x and can be modeled as:

$$U(x) = 1 - \frac{1}{1 + e^{-\alpha(x-\beta)}}. \tag{1}$$

In their definition of LUIF, the authors explain that the QoS received in the distant past has less significant impact on the users' overall long time irritation then the QoS received recently. They use an exponentially weighted moving average (EWMA) to maintain continuous measure of the SUIFs for each user. By letting the stored LUIF be $U(\kappa_{n-1})$, the LUIF to be computed be $U(\kappa_n)$, and the current SUIF $U(x_i)$ be computed proportionally to the time hysteresis outage probability, they used

$$\kappa_n = \rho * \kappa_{n-1} + (1 - \rho) * U(x_i) \tag{2}$$

where ρ is the weight assigned to the cumulative SUIF and κ_n denotes the random variable used to measure the LUIF at the nth request using Eq. (1). We use Eq. (2) in our own work, and we define the variable x in "The proposed MAC scheme".

The related work in regards to resource allocation in satellite networks will be discussed in "Simulation results and discussion", in relation to the proposed scheme's results.

The proposed scheme
CAC based on long term user irritation
As mentioned above, the authors use both SUIF and LUIF in [8]. We argue that CAC decisions *should not be based on SUIF*, especially when taking pricing into account (which [8] does not). As it will be explained in the rest of this section in more detail, our view is that unnecessary degradation of users should not be implemented if there is no revenue gain; however, the provider cannot ignore a profitable policy to avoid an increase in the user's SUIF. This argument is supported by the well-known fact that in wireless networks call dropping creates significantly larger user annoyance than call blocking, i.e., users are more irritated when their ongoing call ends or has a very poor quality than they are when they do not manage to initiate a call. For this reason, and especially in the case that call blocking can lead to a revenue increase, a provider should not take SUIF into account for call admission control decisions, whereas SUIF should be taken into account for medium access control decisions (associated with call dropping). We incorporate SUIF in our MAC scheme presented in "The proposed MAC scheme".

On the other hand, LUIF is of significantly larger importance; a large LUIF could lead a user to drop its contract with a provider. Therefore, LUIF needs to be incorporated into the CAC decisions. This, however, cannot be done for GEO satellites in the manner utilized in [8] for cellular networks, i.e., to use LUIF to make preemption decisions, because this is practically impossible due to propagation delays that would significantly delay the notifications to all users involved in the preemption.

After explaining how our CAC scheme takes provider profit into account, our new proposal on how LUIF can be incorporated into the scheme will be discussed.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 8 of 18

Regarding the revenue consideration: in the case that the admission of a new call (and the subsequent increase in bandwidth utilization) can only be made with the degradation of a higher-paying customer who enjoys higher QoS, the CAC module should compute whether this is a profitable decision. The term "degradation" refers to a "mode" being downgraded to the immediate next "mode" within one of the three groups noted in "Related work" and Table 1. The specific division in groups is only used as an example; any division which takes into account the trace statistics can be used without any qualitative change in our results, as it was shown from our simulations.

For the computation of the profit of a possible degradation, "revenue weights" are computed and assigned for each one of the ten "modes", thereby differentiating them into different service classes. To define what the revenue weights should be, based on network congestion and the type of users present in the network at any given time, we use dynamic pricing. Based on the formula for the demand function from [31], which is implemented for different priority users and fits our system's assumptions with HQ and LQ users, we derive:

$$p = p_o + p_o * \frac{\sqrt{-4\ln(q)}}{2}, \quad p_h \geq p_o \tag{3}$$

where p_o is the price for a low quality user, p_h is the price charged to high quality users and q is the percentage of high quality users who accept dynamic pricing (i.e., they do not accept degradation and are willing to pay more for their calls during network congestion periods). Without loss of generality, the revenue weight has been set equal to 1 for the mode with the lowest bandwidth requirements. With these p_o values, the values of p_h for all the other modes, using Eq. (3), are dynamically calculated. The dynamic calculation is based on the value of q in every time interval of T = 0.5 s, equal to the frame duration. The initial revenue weights, shown in Table 1, are calculated based on the q values for each mode; these values have been selected indicatively, based on the rationale that the modes with the highest bandwidth requirements will be the ones with the least "loyal" users (users who are willing to pay more in order to keep transmitting at a high rate). Hence, for example, Table 1 shows that of the 10% of MPEG-4 users who initially choose, on average, the "mode" Tonight Show with Commercials LQ, 40% are willing to pay more in case of network congestion, in order to keep transmitting at this "mode's" rate. Users who accept degradation are degraded once. Still, q varies at any given time, depending on the traffic mix of the moment. This creates the need of the dynamic calculation of p_h with Eq. (3). Simulations have also been conducted for other values of q and other percentages of "mode" selection (i.e., not 10% for each "mode"). The change in the values had no qualitative influence on our results.

Our CAC scheme uses the traffic model presented in "Related work" to *precompute or compute online* (if a non-precomputed scenario occurs) a number of traffic scenarios. To the best of our knowledge, video traffic prediction for guaranteeing user QoE has not been proposed in the satellite literature.

The current revenue R is computed as:

$$R = \sum_i N_i * W_i \tag{4}$$

where N_i is the total number of video users of "mode" *i*, and W_i is the revenue from each user of "mode" *i*.

The logic of the CAC algorithm is that, when a new video user arrives, the system first checks, with the use of our hybrid video traffic model, whether it can be accommodated in terms of the total bandwidth which will be needed when the user is multiplexed with the existing users in the system.

If this is not possible, the algorithm retrieves the new user's LUIF, which has already been computed from the user's last session in the network (on the contrary, in such a case in [7] the algorithm attempted immediately to degrade the user, without considering user satisfaction). If the LUIF of the new user is larger than the mean LUIF of all users currently in the system, the CAC algorithm will degrade as many users as needed, from those users with LUIF smaller than the mean, in order to ensure that there is enough bandwidth for the new user to be accepted. Of course, only users who accept degradation, based on their contracts with the provider, can be degraded (i.e., users belonging in the 1-q percentage of their "mode"). The new user is rejected only if the CAC computes that, even if all possible users are degraded, there will not be enough bandwidth to accommodate the new user.

If, however, the LUIF of the new user is smaller than (or equal to) the mean LUIF of all users currently in the system, then the CAC algorithm attempts to degrade the new user. The rationale behind this decision is that the arrival of a non-irritated new user should cause the minimum possible number of degradations to users who are already in the system, therefore it is preferable that the new user is accepted with degradation.

If after the degradation of the new user the acceptance of the call is still not possible, the CAC scheme will not degrade a higher priority user, but it will check all possibilities of degrading users of the same or lesser priority of the new call in order to accommodate it. However, the new call will be accommodated *only if its acceptance will lead to higher revenue*; otherwise, even if the total bandwidth that will be used with the acceptance of the new call is larger than the bandwidth previous used, there is no reason to degrade a significant number of users and cause their irritation if the provider will receive no extra revenue. In the case that the new call does not accept any degradation, the attempt to degrade lesser or equal priority users who are already in the system is still made, and the new call is again accepted only if it leads to higher revenue.

Additionally to the incorporation of LUIF into our scheme, we make one more change to our work in [7], regarding Eq. (1). It needs to be pointed out that *β* determines when the utility decreases, while *α* determines the user's sensitivity to the increase/decrease of its irritation. In this study, common QoS requirements for all video users are considered: the two QoS metrics set in this work are that *the video packet dropping probability should not surpass the 0.1% upper bound and that the mean video packet delay should not surpass the 0.6 s upper bound.* A packet is dropped by the terminal if it is not transmitted within 0.6 s.

The upper bound on the mean video packet delay has been selected to be especially strict considering that, for each possible failure of our prediction due to underassignment, the respective packets which would have to wait for a new assignment will have a minimum video packet delay of 0.54 s. The upper bound on the video packet dropping has been selected not only to be strict (in [32] the upper bound is defined as 1%) but

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 10 of 18

also to account for the fact that if a video packet loss corresponds to an I frame, then the "glitch" in the video will propagate to the rest of the video frames in the GoP. Therefore, we chose a much lower upper bound to ensure that the achieved QoS will definitely translate to acceptable video quality.

Given that the QoS requirements are common for all video users, we allow β to be equal to zero, as in [7, 8]. However, contrary to [7, 8] where α is considered to be a constant, in this work *we exploit our video traffic model's ability to distinguish between high and low activity scenes.* More specifically, different values of a are used for high activity scenes (smaller α, i.e., larger irritation) than for lower activity scenes. Also, an even smaller value of a is used for the scenes that immediately follow a change in activity (i.e., the first high activity scene after a low→high transition, and the first low activity scene after a high→low transition); the reason is that after such a transition, the change in video content is the largest, therefore video packet losses are the most costly in terms of user irritation.

The mean value of α is set to {0.1, 0.3 and 0.9} for the groups which have three "modes" and {0.1, 0.3, 0.6, 0.9} for the group which has four "modes". For each mean value of α, a_{HAS} for high activity scenes is set to be 20% smaller than the mean, a_{LAS} for low activity scenes to be 20% larger than the mean and a_{SC} for the first scene after a change in activity to be half the mean value. Our results have shown that the use of other values of α did not alter the nature of our conclusions.

The proposed MAC scheme

The proposed MAC scheme is identical to that of [7]. It is presented here in order to facilitate the readers' understanding of the proposed framework. Both SUIF and LUIF are used in the decisions of the MAC protocol.

As in [33] our proposed satellite MAC scheme is based on a multi-frequency time division multiple access (MF-TDMA) approach, according to which a carrier is divided in timeslots (grouped in frames and superframes). MF-TDMA was chosen since it is an attractive framing scheme for satellite uplinks with low power terminals and MF-TDMA schemes are capable of providing efficient bandwidth utilization [10, 33].

The NCC allocates to each active terminal a set of timeslots, each characterized by a frequency, bandwidth, start time and duration time. Using the highly accurate modeling of multiplexed video traffic, the NCC must run a real-time simulation, to *predict* the traffic volume from video sources. Hence, based on the "mode" declared by the terminals at call establishment, the NCC does not need to wait for a request from the terminals every channel frame (which would arrive with a delay of more than ten channel frames, due to the propagation delay). Instead, it can start allocating resources to the video terminals, and subtracting the estimated used slots from the total number of slots in the system to find the number of free slots. Free slots are allocated based on the user irritation factor, in a manner that will be explained below. The free capacity distribution performed by the protocol brings the end-to-end delay performance at low loads close to that obtained with random access protocols, while the demand-based bandwidth allocation at the beginning of each frame guarantees the protocol's stability, robustness and efficient utilization of transmission bandwidth at high loads.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 11 of 18

Active terminals send a "corrective" request every superframe (defined in our work as equal to 11 channel frames, to account for the propagation delay) to correct any mistakes (due to either overassignment or underassignment of slots) of the models produced at the NCC via online simulation. Terminals send their capacity requests embedded in the header of their packets.

The cross-layer cooperation between the MAC and CAC module is envisaged similarly to [34]; the admission control is performed for a new connection and it may either result on the creation of a new MAC connection, or to the modification of an existent one (call degradation).

Two definitions of SUIF are used in our scheme:

First SUIF definition Video packet transmission delay leads to packet dropping, which in turn leads to user irritation. For this reason, if $x_{1,J}$ denotes the random variable representing the SUIF, normalized with the best possible value being 0 (representing zero delay and jitter) and the worst being 1, then

$$x_{1,J} = \tau * P_{drop} \tag{5}$$

where P_{drop} is the mean packet dropping probability and $\tau < 1$ is the quantitative factor associated with irritation suffered due to a new or handoff call.

Second SUIF definition Typical video encoders use a fixed group-of-pictures (GOP) pattern when compressing a video sequence. The decoding of an I frame in a typical MPEG-4 trace is independent of other video frames. The decoding of P frames depends on the successful decoding of the I frame. The decoding of B frames depends on the successful decoding of I and P frames. Therefore, the successful transmission of an I frame is of paramount importance, while the transmission of P and B frames is important but not as crucial as that of the I frame. Hence, we set

$$x_{1,J} = \tau * P_{THRU\_P,B,GOP} \tag{6}$$

where

$$P_{THRU\_P,B,GOP} = \begin{cases} \frac{P_{TRANS\_P.B}}{P_{GEN\,P,B}}, & if\ P_{drop_{I\_FRAME}} \leq 0.1\% \\ 1, & if\ P_{drop_{I\_FRAME}} 0.1\% \end{cases}$$

The above definition denotes that, if the video packet dropping for the I frame within a GOP exceeds the 0.1% threshold, the transmission of P and B frames is of minor importance, since the basic information from the I frame is missing. If, on the other hand, the basic information from the I frame has been transmitted, then we need to quantify the additional information that manages to be transmitted and without which GOP distortion and user irritation will increase. This is implemented via the calculation, in Eq. (6), of the ratio of the transmitted versus the total generated packets of P and B frames within a GOP.

The rest of the bandwidth is distributed by comparing the high quality users of each group in terms of their LUIF, and then continuing with a comparison among the medium and low quality "modes", respectively. The remaining bandwidth is, each time, allocated to the user with the highest LUIF, in each quality. The bandwidth distribution continues to the remaining users until no more bandwidth is available.

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 12 of 18

In order to quantify the above description of our protocol, we denote by N the number of information slots in the system, L the number of active video stations, $Di(s)$ the amount of bandwidth that the NCC estimates as needed by the ith active terminal at the start of frame s, and $Ai(s)$ the amount of bandwidth that the NCC assigns to the ith active terminal at the start of frame s. If $N - \sum_i^L Di(s) < 0$, i.e., if the NCC estimates that there will be no amount of bandwidth left after all video terminals are granted bandwidth equal to their predicted demands, then the use of the following equation ensures the fair sharing of the available bandwidth resources to all active video terminals:

$$Ai(s) = N * Di(s) / \sum_i^L Di(s). \tag{7}$$

It should be noted here that, although generally $Di(s)$ refers to the estimation made by the NCC of each video terminal's upcoming bandwidth requirements, it also refers, every 11 frames, to the "corrective" request sent by the video terminals every superframe. With the use of Eq. (7) for bandwidth allocation, our protocol also guarantees that, in the case of traffic overload, all users experience equal video packet dropping probability.

If $N - \sum_i^L Di(s) > 0$, i.e., if the NCC estimates that there will be an amount of bandwidth left after all video terminals are granted bandwidth equal to their predicted demands, then the allocation based on user irritation takes place. The major reason behind this policy is fairness: we want to alleviate any mistakes caused by underassignment to a terminal due to a mistaken estimation of its actual larger bandwidth demands. This policy also serves in the case where the NCC has made slight overestimations for all users, as it helps to maximize the number of satisfied users.

Simulation results and discussion

We use computer simulations (the code is written in C) to study the performance of FPRRA with the use of both SUIF definitions. Each simulation point is the result of an average of 100 independent runs (Monte-Carlo simulation), each simulating three hours of network operation. All our results have been derived for 95% t-confidence intervals (constructed in the usual way [29]). Connection lifetimes are exponentially distributed with mean value equal to 180 s. This value has been chosen based on the video marketing survey and business trend report presented in [35], according to which videos under one minute enjoy 80% viewer retention up to the 30-s mark, videos 2–3 min in length enjoy about 65% retention until their 50% mark and 5–10 min videos enjoy close to 50% viewer retention until their 50% mark. By averaging (length × retention)/2, we get an average video viewing time of about 180 s.

The system parameters are taken from [33]: frame duration equal to 26.5 ms, 4 carriers, 128 slots/frame/carrier, 53 bytes/slot, 8 Mbps system global rate. These parameters comply with the relevant ITU-R recommendation [36] and ETSI standard [37, 38].

In EWMA mechanisms, values of ρ between 0.7 and 0.8 are generally chosen [39] (the value λ, defined there as being between 0.2 and 0.3 is equal to $1 - \rho$ for our scheme) although these values are arbitrary and depend on the problem that is being analyzed and, respectively, on the weight that needs to be placed upon current values in

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 13 of 18

comparison to older ones. We have used both the 0.7 and 0.8 values, as well as small values (0.3, 0.4) for ρ and found no qualitative difference in the results. The results presented below have been derived with a value of ρ equal to 0.8. Also, we have experimented with values of τ ranging from 0.3 to 0.8, again without qualitative difference in our results. The results presented below have been derived with τ equal to 0.5.

We compare FPRRA with four other efficient schemes, from [10–12] and from an "ideal" framework, in which the NCC would "magically" know, without any information exchange exactly what the video terminals' bandwidth demands for the next video frame would be. Therefore, no contention is necessary among video terminals. *These schemes do not take into account either pricing or user irritation, therefore they have the advantage over FPRRA that their only goal is the maximization of resource utilization.* Still, as it will be shown from our results, FPRRA outperforms all schemes except the "ideal" framework.

Figure 4 presents our simulation results for the average video packet dropping metric versus the system utilization. Utilization indicates the traffic load normalized to the uplink capacity, e.g., a traffic load equal to 20% represents 20% of the 8 Mbps uplink capacity, i.e., 1.6 Mbps system throughput. As shown in the Figure, FPRRA clearly outperforms the other three protocols from the literature, and is outperformed only by the "ideal" framework. FPRRA can handle up to 59% system load (for the 1st SUIF definition) while at the same time satisfying the strict QoS requirement of maximum video packet dropping equal to 0.1%; the respective maximum system load which the "ideal" framework can handle is 73%, while SRMA-DF [12] achieves only a 23% maximum
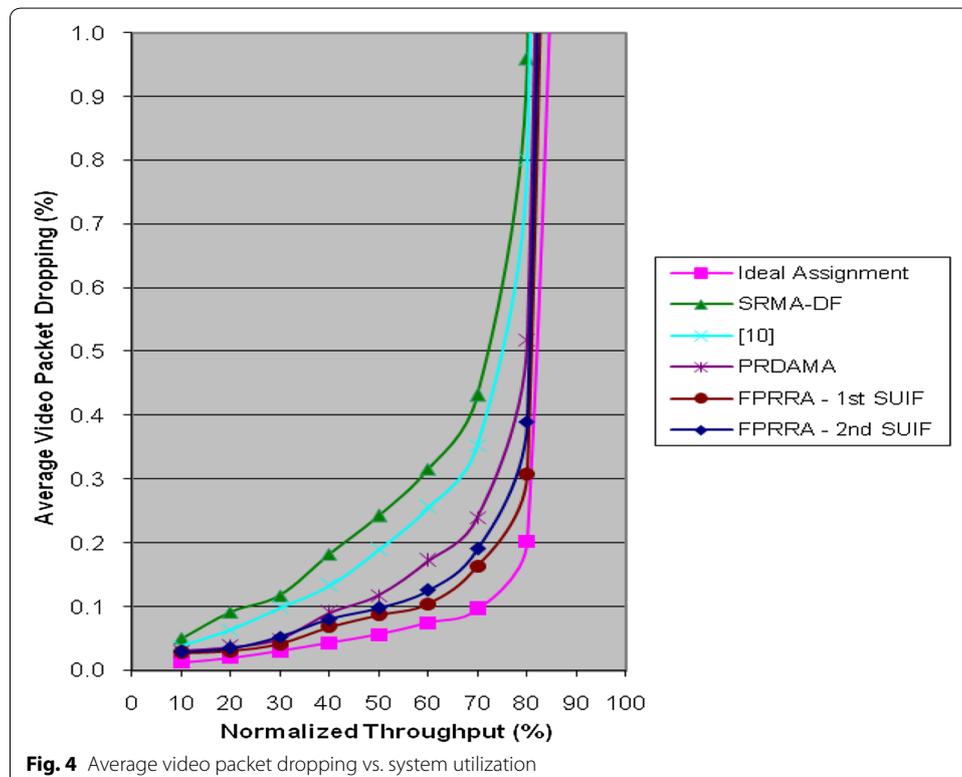

Fig. 4 Average video packet dropping vs. system utilization

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 14 of 18

throughput, [10] achieves a 32% maximum throughput and PRDAMA [11] a 42% maximum throughput for the same QoS requirement.

The reason for the significantly better performance of FPRRA is the use of the new accurate video traffic model for both CAC and MAC purposes, as explained in "CAC based on long term user irritation" and "The proposed MAC scheme". The individual weaknesses of the other three protocols in comparison to a protocol using accurate video traffic prediction have been outlined in [3]. What needs to be emphasized, however, is that the results of FPRRA in this work are worse than the respective ones in [3]. There are two reasons for this.

The first reason that FPRRA cannot achieve a higher throughput is the very high burstiness of video traffic (much higher than the burstiness of videoconference traffic, used in our prior work).

The second reason for FPRRA's inability to achieve a higher throughput is the use of LUIF, both in its CAC component and in its MAC scheme. The use of LUIF in the CAC scheme can lead to the degradation of more than one user with low LUIF, in order to accommodate the new video call. If LUIF wasn't considered, it is possible that the new call would have had to be the only one to be degraded or it would have been rejected, if it could not lead to an increase in provider profit. Both of these outcomes (degradation/ rejection of the new call) would in most cases lead the system to achieve higher throughput than the case when the new user is accepted but more than one existing users are degraded. Also, the use of LUIF in the MAC scheme can lead to prioritizing for transmission users which have a higher LUIF but their transmission deadline is not approaching, whereas other users with lower LUIF and imminent deadlines will have to wait. This leads to suboptimal channel utilization in order to preserve fairness in terms of user satisfaction. Hence, despite its importance, the use of LUIF comes with a "user satisfaction vs. throughput" tradeoff.

In [7] it was found that the use of the 2nd SUIF definition leads to a fluctuation, in comparison to the results with the use of the 1st SUIF definition; in other cases the results were marginally lower with the 1st definition, and in other cases they were lower when the 2nd is used. In the present study the results with the use of the 2nd SUIF definition are constantly worse. The reason is that the 2nd SUIF definition increases user "sensitivity" (i.e., irritation) in the cases where the loss of information is concentrated in specific GoPs, whereas the 1st SUIF definition "triggers" user irritation when packet dropping occurs anywhere in the video frames' transmission. Due to the use of much smaller α values in high activity scenes and in scene activity changes, the different behavior of the system with the use of the 2nd SUIF is accentuated.

The results achieved with FPRRA for the 1st SUIF (handling up to 59% system load) also clearly excel against the contention-based protocol CRDSA [40] which achieves a peak throughput of 52%. Still, a full comparison cannot be made, since CRDSA only contains a MAC component. Also, in comparison with the efficient CAC scheme presented in [41], FPRRA achieves much better results (lower packet loss percentage for much burstier sources than the ones used in that paper), and additionally takes into account user satisfaction and provider revenue, whereas the scheme in [41] only bases its decisions on the capacity needed to accommodate the call. Also, [41] assumes that each

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 15 of 18

video source's rate is divided into a fixed number of discrete bandwidth levels, which for multiple sources, possibly generating context on-the-fly, is impractical. The work in [42] supported the use of active measurements to allow bandwidth adaptation and consequent tracking of the chosen performance metrics. However, in the case of bursty multimedia traffic, this approach is inefficient due to the significant fluctuations in the sources' rate. In the case of such fluctuations, the use of active measurements can lead to considerable bandwidth overallocation or underallocation before the next round of measurements gives the system the opportunity to react.

Figure 5 presents our simulation results for the average video packet delay versus the system utilization. The results are generally similar in nature with those of Fig. 4, including those concerning the comparison of the two SUIF definitions. Also, in Fig. 5, the confidence intervals for FPRRA for both SUIF definitions are presented. The graphs connecting all the low confidence interval (LCI) and high confidence interval (HCI) values, respectively, show that the qualitative differences between all the compared schemes remain unaltered even for the lowest and highest values of LCI and HCI. In this Figure, FPRRA1 stands for FPRRA-1st SUIF.

In the results presented in Fig. 6 we use Jain's fairness index [43] in order to evaluate the system behavior under each of the two SUIF definitions. Fairness is studied in terms of the video packet dropping encountered by individual video streams when using each SUIF definition. From the results it is clear once again that the 2nd SUIF definition leads to a larger LUIF and hence to decreased fairness; on the other hand, the 2nd SUIF definition is of higher practical value, because it takes into account the interdependence among video frames.

Finally, we compared FPRRA against our prior work in [7] in terms of LUIF. The incorporation of LUIF into FPRRA's CAC decisions, combined with our new model's accuracy, was shown to decrease the average Long Term User Irritation by 27%, over all of our experiments.
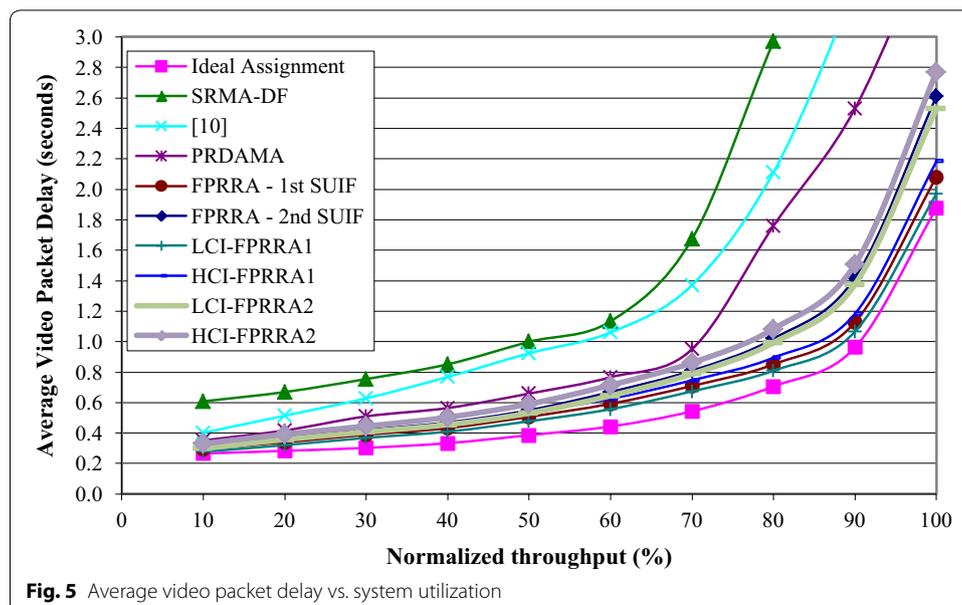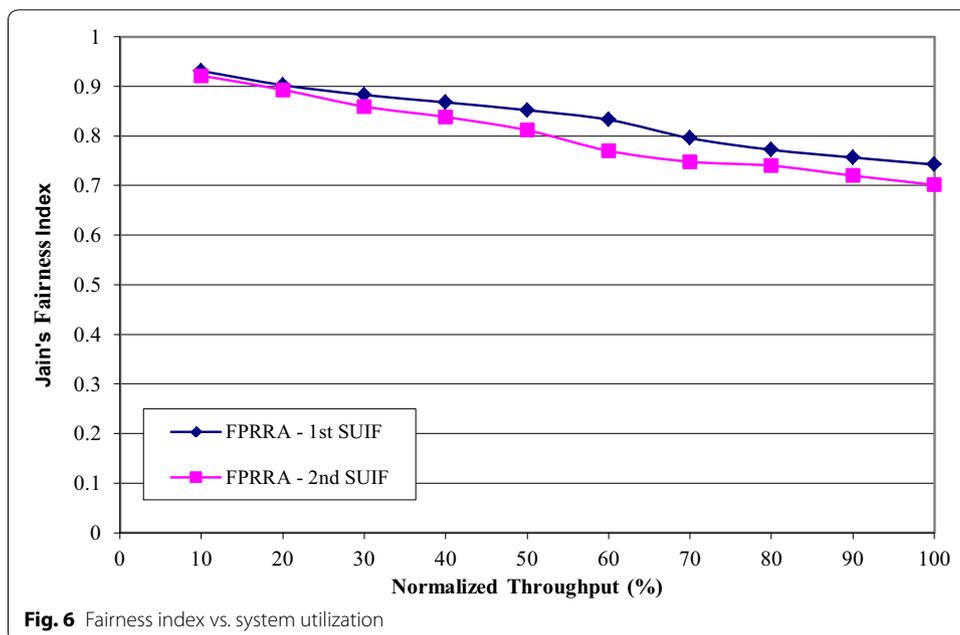

Fig. 5 Average video packet delay vs. system utilization

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 16 of 18

Fig. 6 Fairness index vs. system utilization

Conclusions

In this paper we have proposed, for the first time in the relevant literature to the best of our knowledge, the use of video scene identification and classification to improve user QoE over GEO satellite links.

We introduced the notion of long-term user satisfaction into the call admission control scheme, and mapped our model's ability to distinguish between high and low activity scenes into QoE. This led our scheme, which makes decisions based on available bandwidth, user satisfaction and the possible revenue for the provider, to be able to handle the very bursty video traffic and to outperform other schemes from the literature in all of the QoS and QoE metrics used in our study.

The proposed scheme, additionally to its superior results, has the advantages of: (a) using a relatively simple video model towards video scene identification and classification, (b) adhering to the general directions set in the ETSI standard [44], according to which the streaming class needs to be supported for video admission control and the scheduling process needs to be driven primarily by the connection Quality of Service parameters.

In future work we intend to study the possible extension/combination of our scheme with recent proposals to exploit successive interference cancellation schemes, in order to solve packet collision issues [45]. We also intend, in line with our view towards cross-layer extensions of our work, to study how our scheme can be combined with the DVB-S2X extension of the DVB-S2 specification that provides additional technologies and features for the core applications of DVB-S2. As noted in [46], the DVB-S2X has been introduced at the same time as the new high efficiency video coding (HEVC) scheme and it is expected that new satellite DTH receivers will combine these two technologies to make the delivery of ultra high definition services more efficient. For this reason

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 17 of 18

we also intend to evaluate our scheme with HEVC/H.265 video traces, as HEVC is fast becoming widely adopted. We believe that the combination of our scheme with DVB-S2X will be a significant step towards enabling its practical implementation.

Authors' contributions
PK proposed the use of video scene identification and classification for traffic modeling at the scene level and the utilization of its relation to user satisfaction in order to improve system utilization and user QoE in GEO satellite networks. PK proposed and implemented the CAC algorithm using Long Term User Irritation to make acceptance/rejection decisions. IS and AL implemented the video traffic modeling module, analyzed the results of the proposed scheme and helped to draft the manuscript, together with PK. All authors read and approved the final manuscript.

Author details
[1] School of Engineering and Information Technology, Murdoch University, Science and Computing Building 245, SC1.012, 90 South Street, Murdoch, WA 6150, Australia. [2] Global Virtual Engineering, Cisco Systems, Marousi, Greece. [3] Department of Electrical Engineering, University of Southern California, Los Angeles, USA.

Competing interests
The authors declare that they have no competing interests.

References
1. Gotta A, Luglio M, Roseti C (2014) A TCP/IP satellite infrastructure for sensing operations in emergency contexts. Comput Netw 60:147–159
2. NetWorld2020–SatCom Working Group (2014) The role of satellites in 5G. https://www.networld2020.eu/wp-content/uploads/2014/02/SatCom-in-5G_v5.pdf. Accessed 21 Mar 2017
3. Koutsakis P (2011) Using traffic prediction and estimation of provider revenue for a joint GEO satellite MAC/CAC scheme. Wirel Netw 17:797–815
4. Wittig M (2000) Satellite onboard processing for multimedia applications. IEEE Commun Mag 38:134–140
5. Le-Ngoc T et al (2003) Interactive multimedia satellite access communications. IEEE Commun Mag 41:78–85
6. Lazaris A, Koutsakis A, Paterakis M (2008) A new model for video traffic originating from multiplexed MPEG-4 videoconference streams. Perform Eval 65:51–70
7. Stamos C, Vasileiadou D, Koutsandria G, Spanou I, Vlachaki A, Lazaris A, Koutsakis P (2012) User-satisfaction based resource allocation for GEO satellites. In: Paper presented at the IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), San Francisco
8. Pal S, Chatterjee M, Das SK (2005) A two-level resource management scheme in wireless networks based on user-satisfaction. ACM Mobile Comput Commun Rev 9:4–14
9. Spanou I, Lazaris A, Koutsakis P (2013) Scene change detection-based discrete autoregressive modeling for MPEG-4 video traffic. In: Paper presented at the IEEE international conference on communications (ICC), Budapest
10. Iuoras A et al (1999) Quality of service-oriented protocols for resource management in packet-switched satellites. Int J Satell Commun 17:129–141
11. Jiang Z, Leung VCM (2003) A predictive demand assignment multiple access protocol for internet access over broadband satellite networks. Int J Satell Commun Netw 21:451–467
12. Yum TS, Wong EWM (1989) The scheduled-retransmission (SRMA) protocol for packet satellite communications. IEEE Trans Inf Theory 35:1319–1324
13. Maglaris B (1988) Performance models of statistical multiplexing in packet video communications. IEEE Trans Commun 36:834–844
14. Liu D, Sara E, Sun W (2001) Nested auto-regressive processes for mpeg-encoded video traffic modeling. IEEE Trans Circuits Syst Video Technol 11:169–183
15. Krunz M, Tripathi SK (1997) On the characterization of VBR MPEG streams. ACM Sigmetrics Perform Eval Rev 25:192–202
16. Dawood AM, Ghanbari M (1999) Content-based MPEG video traffic modeling. IEEE Trans Multimed 1:77–87
17. Melamed B, Pendarakis DE (1998) Modeling full-length VBR video using Markov-renewal modulated TES models. IEEE J Sel Areas Commun 16:638–649
18. Chandra K, Reibman AR (1999) Modeling one- and two-layer variable bit rate video. IEEE/ACM Trans Netw 7:398–413
19. Heyman DP (1997) The GBAR source model for VBR videoconferences. IEEE/ACM Trans Netw 5:554–560
20. Frey M, Ngyuyen-Quang S (2000) A gamma-based framework for modeling variable-rate video sources: the GOP GBAR model. IEEE/ACM Trans Netw 8:710–719
21. Dai M, Zhang Y, Loguinov D (2009) A unified traffic model for MPEG-4 and H.264 video traces. IEEE Trans Multimed 11:1010–1023
22. Bhattacharya A (2003) Prediction of MPEG-coded video source traffic using recurrent neural networks. IEEE Trans Signal Process 51:2177–2190
23. Huang CL, Liao BY (2001) A robust scene-change detection method for video segmentation. IEEE Trans Circuits Syst Video Technol 11:1281–1288

Koutsakis *et al. Hum. Cent. Comput. Inf. Sci.* (2017) 7:15

Page 18 of 18

24. Lelescu D, Schonfeld D (2003) Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream. IEEE Trans Multimed 5:106–117
25. Chiruvolu G et al (1998) A scene-based generalized markov chain model for VBR video traffic. In: Paper presented at the IEEE international conference on communications (ICC), Atlanta
26. Yoo SJ (2002) Efficient traffic prediction scheme for real-time VBR MPEG video transmission over high-speed networks. IEEE Trans Broadcast 48:10–18
27. Radke RJ et al (2005) Image change detection algorithms: a systematic survey. IEEE Trans Image Process 14:294–307
28. Seeling P, Reisslein M, Kulapala B (2004) Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial. IEEE Commun Surv Tutor 6:58–78
29. Law AM, Kelton WD (1991) Simulation modeling & analysis, 2nd edn. McGraw Hill, New York City
30. Burnham KP, Anderson DR (2002) Model selection and multi-model inference. Springer, New York
31. Yaipairoj S, Harmantzis F (2004) Dynamic pricing with alternatives for mobile networks. In: Paper presented at the IEEE wireless communications and networking conference (WCNC), Atlanta
32. Lewis C, Pickavance S (2006) Implementing Quality of Service over CISCO MPLS VPNs. CISCO Press, Indianapolis. http://www.ciscopress.com/articles/article.asp?p=471096&seqNum=6. Accessed 14 Dec 2016
33. Chiti F, Fantacci R, Marangoni F (2005) Advanced dynamic resource allocation schemes for satellite systems. In: Paper presented at the IEEE international conference on communications (ICC), Seoul
34. Melhus I et al (2008) Cross-layer optimization in the next-generation broadband satellite systems. In: Paper presented at the international communications satellite systems conference (ICSSC), San Diego
35. Singlegrain.com (2014) Just the stats: the science of video engagement. https://www.singlegrain.com/video-marketing/just-stats-science-video-engagement/. Accessed 21 Mar 2017
36. International Telecommunication Union (2012) Cross-layer QoS provisioning in IP-based hybrid satellite-terrestrial networks, Recommendation ITU-R S.1897, January 2012
37. European Telecommunications Standards Institute (2005) Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia; Transparent Satellite Star-A (TSS-A); DVB-S and DVB-RCS for Transparent Satellites; Sub-family 1 (TSS-A1), May 2005
38. European Telecommunications Standards Institute (2006), Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Regenerative Satellite Mesh-B (RSM-B); DVB-S/DVB-RCS Family for Regenerative Satellites; Part 1: System Overview, October 2006
39. Hunter JS (1986) The exponentially weighted moving average. J Qual Technol 18:203–210
40. Casini E, De Gaudenzi R, del Rio Herrero O (2007) Contention resolution diversity slotted ALOHA (CRDSA): an enhanced random access scheme for satellite access packet networks. IEEE Trans Wirel Commun 6:1408–1419
41. De Rango F et al (2008) Call admission control for aggregate MPEG-2 traffic over multimedia geo-satellite networks. IEEE Trans Broadcast 54:612–622
42. Marchese M, Mongelli M (2006) On-line bandwidth control for quality of service mapping over satellite independent service access points. Comput Netw 50:2088–2111
43. Jain R (1991) The art of computer systems performance analysis. Wiley, New York
44. European Telecommunications Standards Institute (2015), Satellite Earth Stations and Systems (SES); Family SL Satellite Radio Interface (Release 1); Part 1: General Specifications; Sub-part 3: Satellite Radio Interface Overview, October 2015
45. De Gaudenzi R et al (in press) Random access schemes for satellite networks, from VSAT to M2M: a survey. Int J Satell Commun Netw. doi:10.1002/sat.1204/full
46. Morello A, Migone M (2015) DVB-S2X: the new extensions to the second generation DVB satellite standard DVB-S2. Int J Satell Commun Netw 34:323–325