Semantic-Aware Radio Resource Scheduling for Video Streaming in Mobile Packet Networks

V. Donini, F. Lironi, C. Masseroni, R. Trivisonno

Abstract— The paper presents the Selective Random Discarding (SRD) algorithm, a semantic-aware radio resource management mechanism designed to improve end-to-end performance of multimedia streaming services over mobile and wireless networks. The algorithm comprises two different steps: first, a mechanism to detect critical conditions on the radio interface is introduced; second, if the available radio resources are not sufficient to carry the media stream fulfilling QoS requirements (in terms of packet transfer delay), the less relevant information flow is selectively and partially discarded. Information flows are served on the radio interface according to their relevance at application layer. SRD has been designed referring to MPEG-4 coded media files and to systems based on the Enhanced General Packet Radio Service (EGPRS), the 2.5G packet data service. The impact of the algorithm on end-to-end performance has been evaluated considering different metrics. The average PSNR, the IP packet transfer delay and the probability of stall and re-buffering during stream play-out have been considered.

Keywords— Multimedia Streaming; QoS; Semantic-Awareness; Radio Resource Management; Cross-Layer Design.

I. Introduction

THE need to support multimedia streaming services over mobile networks has considerably determined their recent evolution. Many enhancements have been conceived and introduced in order to fulfill multimedia streaming and real-time QoS requirements. Nevertheless, due to the high (unpredictable) variability over time of the bandwidth required by an audio-video stream and to the (unpredictable as well) time-varying link capacity, an effective and efficient radio resource dimensioning can not always be guaranteed. The attempt to avoid the waste of Radio Resources -RR- (efficient dimensioning), accomplished via optimized allocation algorithms, can lead to insufficient resources during situations where many streaming sessions are active (ineffective dimensioning), thus leading to the inability to provide satisfactory end user QoS perception. Exploiting the novel *Cross-Layer* designing principles, however, Radio Access Network (RAN) functionalities can be improved to overcome this problem. This paper presents a novel "*semantic-aware*" radio resource scheduling algorithm, based on the cross-layer design principle, conceived with the aim to allow a continuous access to a multimedia streaming service without interruption for rebuffering, even in case of temporarily insufficient radio link capacity and with minor degradation of end user QoS perception. The paper is organized as follows: section II A describes the Semantic-Aware principle as well as video codec main characteristics; section II B gives an overview on previous work on this topic; section II C focuses on cross-layer principle which lies behind this scheduling method. In section III A the novel *Selective Random Discarding* (*SRD*) algorithm is presented and designing guidelines for its introduction within 2.5G packet switched data network are examined in section III B. The simulation activity carried out for *SRD* performance evaluation is presented in section IV. Section V concludes the paper.

Valentina Donini, Francesco Lironi, Carlo Masseroni and Riccardo Trivisonno are with SIEMENS Mobile Communications, 20092 Cinisello Balsamo, Milan, Italy (emails: valentina.donini@icn.siemens.it - francesco.lironi@icn.siemens.it - carlo.masseroni@siemens.com - riccardo.trivisonno@siemens.com)

II. Semantic Aware RR Scheduling at RAN

A. Semantic-Aware Scheduling: Basic Principle

Advanced video coders exploit spatial and temporal redundancy of the source data. Intra-frame coding is used to reduce spatial redundancy, whilst temporal redundancy is reduced via inter-frame coding. MPEG-4, the variable bit rate (VBR) video coding expected to be used for 3G wireless network services [1], exploits inter frame coding. Three different kinds of frames (namely VOP – Visual Object Plane) are foreseen: I-frames (pictures coded using information on the picture itself only, i.e. frames coded independently of each other), P-frames (*predicted frames*, coded depending on previous I-frames) and B-frames (*bidirectional frames*, coded depending on previous and following I and P-frames). I, P and B frames differ in size and are arranged in periodic pattern (GOP – Group of visual object Plane) in which one I-frame only and some related P and B frames are present. I-frames carry the most information content; moreover, loosing a single I-frame causes the distortion in the decoding of all P and B frames belonging to the same group. I-frames can be identified as the most critical information that should be treated with high priority when bandwidth requirements can not be fulfilled due to critical conditions on the radio interface. The principle of "semantic-aware" radio resource management lies in the ability to selectively manage the information flow depending on its importance from an application point of view. The introduction of such a scheme in any wireless network can be accomplished via two distinct steps: a) making radio resource manager/scheduler aware of the priority of the data being transmitted and b) defining the semantic-aware algorithm.

B. Cross Layer Design

The introduction of semantic-aware scheduling at the Radio Access Network (RAN) implies the adoption of the cross-layer designing principle. The Medium Access Control (MAC) entity is part of the Data Link Layer: the Radio Resource (RR) scheduler is implemented at that layer where no semantic information is available. In addition, the RR scheduler manages Data Link LL-PDUs (Logical Link Protocol Data Units) which are generated by fragmenting IP packets; generally, application frames may be split over several LL-PDUs and hence it is not possible to identify the application frame being transmitted managing LL-PDUs. Semantic information are available at application layer, where coders are implemented. Cross-layer information flow is required. Moreover, it has to be noticed that at Radio Access Points, where RR scheduler is located, usually only a subset of the protocol stack is present (usually up to layer 2 or 3). In other words, at the Network Element (NE) implementing the RR scheduler, the semantic information is not available, and hence a proper inter-NEs information distribution has to be introduced. Summarizing, semantic aware scheduling implementation in a real system requires two separate steps:

- Cross-layer/NE semantic information flow introduction;
- Semantic-aware RR Scheduling algorithm definition.

C. Previous Work

The concept of semantic-aware scheduling has already been implicitly presented in some recent works addressing the support of streaming services in wireless networks. In [2, 3, 4] acknowledged and unacknowledged RLC protocols are foreseen for high and low importance information flows respectively. In [5] a different transmission deadline is assigned to packets according to the priority of the information flow they belong to. The same principle is also presented in [6], where a different transmission deadline is associated to the different type of frames characterizing an MPEG video stream.

III. A Novel Semantic Aware- Algorithm

The basic goal of the algorithm presented is to select and transmit the exact amount of data which fits with bandwidth availability in order to guarantee end-to-end transfer delay constraints and to allow the play-out of the stream without interruptions. To achieve this goal, the less relevant information flow is partially and selectively discarded in case of critical conditions on the radio interface. The novelty of the proposal lies in the fact that the selective treatment of information is activated only if critical conditions occur; the selective discarding is considered as an extreme measure and therefore can be performed for limited periods of time.

A. SRD Algorithm Description

The basic information units managed by the radio network, defined as Logical Link Protocol Data Units (LL-PDUs), carry fragmented (if necessary) IP packets, which are containing encapsulated MPEG-4/RTP frames. In order to implement semantic-aware algorithms, the Radio Resource Manager/Scheduler (RRM) is expected to be aware of the information content of each LL-PDU. I, P and B LL-PDUs are referred in the following as PDUs containing I, P and B MPEG-4 frames respectively. Only the video component of the streaming flow is considered in this work.

A streaming session is admitted in the packet switched radio network after a preliminary admission control phase. Basically, the RRM accepts the incoming service request, if it estimates that the average bandwidth required by the streaming service (B_R[Kbps]) can be guaranteed on the radio interface. Recent studies [7] have shown that realizing the average required bandwidth allows to fulfill the LL-PDUs Transfer Delay (TD) constraints. The Streaming QoS profile requires that the TD_P[%] percent of LL-PDUs is to be affected by a TD less than TD_R[s]. During the session, variations on the radio link quality lead to a fluctuating available bandwidth. If the average available bandwidth (B_A[Kbps]) estimated by the RRM falls below the required one, i.e. if the condition:

$$B_A < B_R \quad (1)$$

is true, the situation is assumed to be critical. Nevertheless, the SRD algorithm is not immediately activated: condition (1) could become true, while the multimedia flow is demanding less bandwidth than the amount negotiated during the admission control phase. When (1) is true, assuming the available bandwidth equal to B_A onwards, the number of LL-PDUs for which the TD can be kept less then TD_R is computed:

$$N_{TD} = \frac{B_A \cdot TD_R}{PDU_{Size}} \quad (2),$$

where PDU_{Size} [Byte] is the size of LL-PDUs and it is assumed constant. Defining N_{PDU} the number of LL-PDUs stored within the LL-PDU buffer, the SRD algorithm is activated if:

$$N_{PDU} \cdot TD_P > N_{TD} \quad (3).$$

The SRD algorithm scans the content of the LL-PDU buffer and randomly discards P and B LL-PDUs only (no I-LL-PDUs are cancelled). Each P or B LL-PDU is randomly suppressed with Discarding Probability (DP):

$$DP = \Delta \cdot \left(1 - \frac{N_{TD}}{N_{PDU} \cdot TD_P}\right) \quad (4),$$

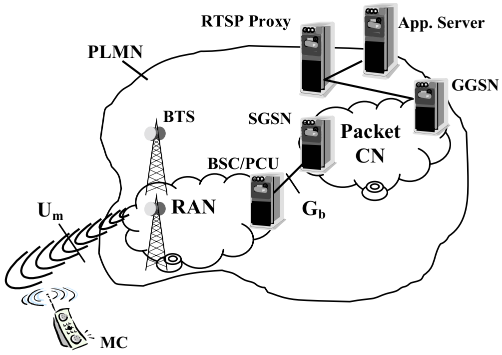

Figure 1. EGPRS Network Basic Architecture.

where Δ(>1) is a tuning parameter. After SRD activation, only (approximately) N_{TD} LL-PDUs remain present in the buffer. Note that the SRD is considered an extreme measure to cope with sudden reductions of the available bandwidth. If permanent radio link worsening takes place, a continuous activation of SRD would lead to unacceptable end user perceived QoS (i.e. the video quality would be permanently too poor). Therefore, a Maximum Activation Timer T_{MAT} is started with SRD: it defines the maximum duration of SRD activation. If T_{MAT} expires (default value 5s), SRD is deactivated. Other countermeasures will be undertaken, depending on the wireless network considered (e.g. packet switched connection reconfiguration).

B. SRD Introduction in 2.5G Network

In order to verify the effectiveness of the proposal in a real system, the introduction of the SRD algorithm within EGPRS network (the 2.5G packet data network) has been designed. The EGPRS network architecture is schematically depicted in Fig 1. As stated in section II B, the real implementation can be accomplished via two steps: spreading semantic information to NEs which require it and implementing the SRD algorithm. Two EGPRS NEs have been affected by the semantic-aware enhancement: the Serving GPRS Support Node (SGSN) and the Base Station Controller/Packet Control Unit (BSC/PCU). The SGSN is the gateway element between the core network and the RAN; IP packets coming from external IP network are routed to SGSN and there fragmented into LL-PDUs. LL-PDUs are transferred to the BSC/PCU via the G_b interface; BCS/PCU manages the radio interface U_m. A *snooping and marking* agent has been introduced in the SGSN: incoming IP packets from the service provider, routed via the GGSN, are snooped into and LL-PDUs created are marked as I, P or B LL-PDUs according to the MPEG-4 frame type carried by the IP packet. The *SRD* algorithm is introduced in the RR manager within BSC/PCU. Fig. 2 depicts the two NEs impacted and the functionalities introduced.

IV. SIMULATION ACTIVITY AND RESULTS

A. Simulation Environment and QoS Metrics Description

The effectiveness of the SRD algorithm has been evaluated in a EGPRS emulation environment. The downstream (from the Internet Service Provider ISP to the Mobile Client MC) of different video clips in real environment has been performed. Several MPEG-4 coded video streams have been considered, with average required bandwidth equal to 32, 49 and 60kbps. An initial buffering time of 15s at MC is considered. Only I and P frames are present in the coded stream. The video downstream has been performed in several real radio scenarios represented by C/I over time traces, taken from on-field measurements. At the beginning of streaming sessions, the admission control and the RR allocation have been performed considering a reference C/I of 17dB, resulting in an average available capacity of 35Kbps per timeslot (TS). The packet switched connection (*Temporary Block Flow TBF*) is initially configured considering the expected average available capacity per timeslot and the expected average required bandwidth of the video stream.

During the session, EGPRS TBF reconfiguration mechanisms are foreseen in order to adapt video stream real requirements with real available bandwidth on the radio interface. Such mechanism is triggered by the LL-PDUs buffer level exceeding the thresholds N_{Rec}=50Kbyte. Reconfigurations, however, are low-reactive mechanisms and often lead to not effective RR exploitation. Hence, they have to be reduced as much as possible.

Different performance metrics have been considered. The improved capability of the radio network to make the streaming QoS constraints respected has been evaluated by monitoring the LL-PDUs transfer delay. End-user perceived QoS has been evaluated considering the occurrence of interruptions and rebuffering in video playout and computing the average *Peak Signal to Noise Ratio* (*PSNR*). The *PSNR* is computed frame by frame, comparing the frame in case of ideal decoding (F) and the actual decoded frame (f). Identifying each frame as a matrix of *NxM* pixels, and defining the *Mean Square Error* (*MSE*) as:

$$MSE(F,f) = \frac{1}{N \cdot M} \cdot \sum_{(i,j)\in(NxM)} \left(F(i,j) - f(i,j)\right)^2$$

the PSNR can be computed as:

$$PSNR(F,f) = 10 \cdot \log \frac{(255)^2}{MSE(F,f)} \quad (5).$$

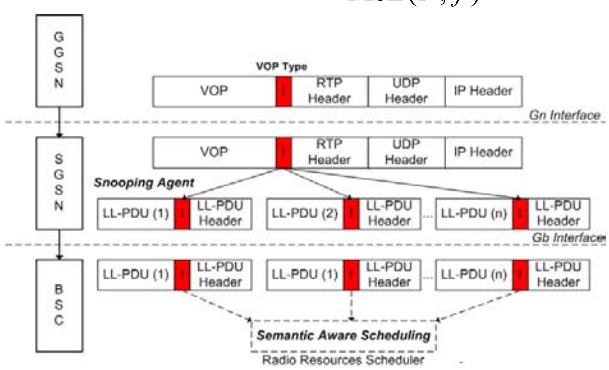

Figure 2. Semantic-Aware in EGPRS Network.

The mapping between the average PSNR and subjective quality metrics, commonly expressed via the "Mean Opinion Scale" (*MOS*), is heuristically performed. In Tab. I *PSNR* ranges and the corresponding *MOS* values are reported.

TABLE I
PSNR-MOS MAPPING

PSNR [db]	MOS
>37	5 (Excellent)
31-37	4 (Good)
25-31	3 (Fair)
20-25	2 (Poor)
<20	1 (Bad)

B. Simulation Results

The Semantic-Aware Scheduling was proved to be effective in scenarios with sudden and temporary reductions of the radio link capacity; on the contrary, no impact is present if the hypothesis of radio conditions assumed in the admission control phase are verified during the whole session. Therefore, this section is focused on three critical real cases in which the average C/I over the session is equal to 13dB, 14dB and 15dB. The first improvement achieved by the introduction of SRD algorithm is related to the capability of the RAN to fulfil QoS profile constraints. Fig. 3 depicts the LL-PDU Transfer Delay (TD) CDF for the 1st scenario (average C/I=13dB) for the 32Kbps and 49Kbps video streams, both in case of SRD enabled and not enabled. The improvement is clearly visible: the 95%ile of the TD shifts from 17s to 13s in case of 32Kbps video clip and from 13s to 8s in case of 49Kbps video clip. Note that during the RR allocation phase, 1 TS is allocated to the TBF for the 32Kbps clip and 2 TSs are allocated to the TBF for the 49Kbps clip (this explains the lower TD values for the 49Kbps case). Obviously, this is a pathological case where radio condition are too poor to support a streaming session with a typical QoS profile (the 95%ile of the TD is expected to be of about 2s); nevertheless, it is suitable to understand SRD functioning principle. Fig. 4 depicts the LL-PDU buffer level for the case of 32Kbps clip. Since the very beginning of the session (t<10s) the available bandwidth on the radio interface is insufficient to carry the stream and, hence, the SRD algorithm is activated. Consequently, LL-PDUs are randomly dropped and the buffer level is kept below 10Kbyte, while if SRD is not introduced the level rapidly rises above 20Kbyte. Note that EGPRS TBF reconfiguration mechanisms take place in case of insufficient link capacity: the TBF is reconfigured twice both in case of SRD enabled and not enabled (the first reconfiguration assigns 2 TSs to the TBF, the second one assigns 3 TSs). As the reconfiguration process is related to the LL-PDU buffer level, reconfigurations a) occur faster in case of SRD not enabled, but b) can not be avoided in such critic radio conditions. Reconfigurations occur approximately at t=27s and t=33s in case of SRD not enabled and at t=43s and t=49s in case of SRD enabled. The boxed curves in Fig. 4 show the bandwidth evolution during the session in case of SRD enabled and not enabled. Discontinuities correspond to TBF reconfiguration instants. Consequently, buffer level profile,

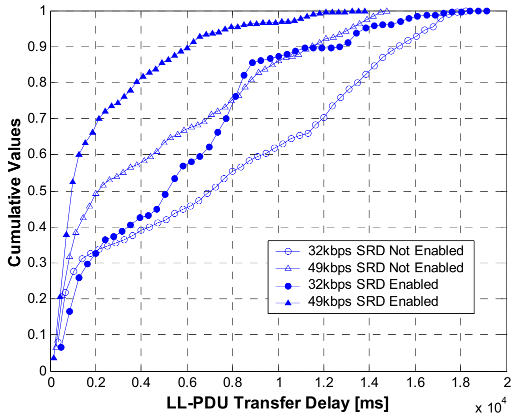

Figure 3. LL-PDU Transfer Delay CDF; 1st Scenario (Average C/I=13dB), 32 and 49Kbps Video Stream

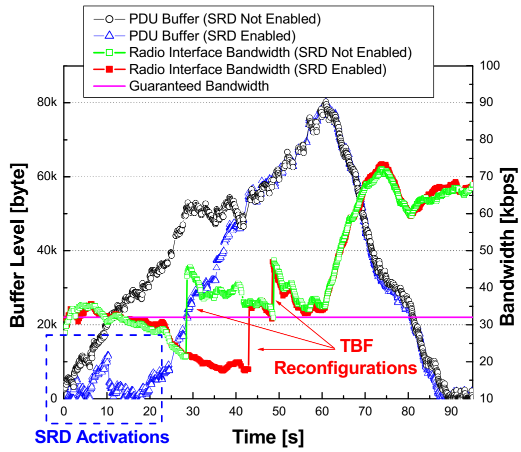

Figure 4. LL-PDU Buffer Level; Available Bandwidth; 1st Scenario (Average C/I=13dB), 32Kbps Video Stream

approximately starting from t=50s, is the same both in case of SRD enabled and not enabled and, anyway, SRD is not activated anymore. Nevertheless, even in this pathological case, the benefit from SRD introduction is evident: considering an initial buffering time of 15s, SRD allow the user to access the whole video clip without interruption for rebuffering. If SRD is not enabled, a stall takes place at instant t=50s, as shown in the protocol tracing depicted in Fig. 5. The quality of the stream worsens, due to the dropping of P-frames, but the average PSNR is anyway kept above 40dB. Table II summarizes results for the three scenarios considered. *SRD* avoids the interruption of the clip in scenarios #1 and #2; in scenario #3, stall does not take place both in case of *SRD* enabled and not enabled; with S*RD,* however, one reconfiguration only (rather than two) occurs. It should also be noticed that *SRD* activation leads to the loss of 7.5%, 15.3% and 11.2% of P-frames in the three cases shown. The average *PSNR* for these cases is anyway greater than 37dB (the corresponding MOS is always equal to 5 – excellent). In the portion of the clip where P-frames dropping occurs, however the PSNR shows fast fluctuations between 10dB and 30dB. The PSNR vs. time for the 1st scenario is shown in Fig. 6. Despite this, the QoS perceived by the user is definitely satisfactory, having a periodic repetition of images with

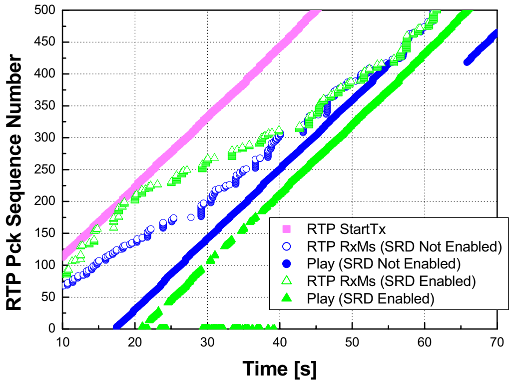

Figure 5. RTP Protocol Tracing; 1st Scenario (Average C/I=13dB), 32Kbps Video Stream

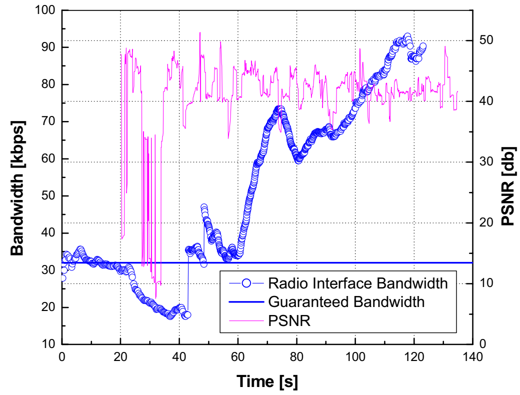

Figure 6. PSNR and Available Bandwidth; 1st Scenario (Average C/I= 13dB), 32Kbps Video Stream

TABLE II
OVERALL SIMULATION RESULTS

		Scenario #1	Scenario #2	Scenario #3
Average C/I [db]		13	13	15
Required Bandwidth [kbps]		32	49	32
Number of Time Slots		1	2	1
Number of Reconfig.	No SRD	2	2	2
	SRD	2	2	1
Number of Stalls	No SRD	1	1	0
	SRD	0	0	0
% P-frame Lost (SRD)		7.5	15.3	11.2
Avg PSNR (SRD) [db]		40.62	38.96	41.79

Figure 7. Sequence of images played every 100ms during SRD activation PSNR fluctuating between 10dB and 35dB

different level of degradation. An example of sequence of 6 images lasting 600ms is shown in Fig. 7. Impacting improvements have been highlighted with the simulation activity: even in case of very critical radio interface conditions, the SRD introduction either avoids the stall of media clip or reduces the number of TBF reconfigurations, with no significant worsening of the user perceived QoS of the video.

V. CONCLUSIONS

A novel semantic aware radio resource scheduling algorithm has been proposed. The SRD algorithm has been designed in detail referring to MPEG-4 multimedia streams and it has been developed for the introduction in EGPRS networks.

SRD, which does not have any impact in case of good radio conditions, has been proved to be effective in case of critical real scenarios. In a number of cases examined, SRD allows to playout the video without interruptions and rebuffering or limiting the number of TBF reconfigurations, with a minor degradation of the end under perceived QoS.

REFERENCES

[1] 3GPP TS 26.234 v5.3.0, Technical Specification, 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects; Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs; (Release 5);

[2] *Fabri, S.N.; Worrall, S.; Sadka, A.; Kondoz, A.;* **“**Real-time video communications over GPRS”; 3G Mobile Communication Technologies, 2000. First International Conference on (IEE Conf. Publ. No. 471) , 27-29 March 2000;

[3] *Fabri, S.N.; Kondoz, A.M.;* **“**Provision of streaming media services over mobile networks”**;** 3G Mobile Communication Technologies, 2001. Second International Conference on (Conf. Publ. No. 477) , 26-28 March 2001;

[4] *Fabri, S.N.; Farrugia, M.; Kondoz, A.M.;* “Provision of scalable streaming audio services over EGPRS”; Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd , Volume: 3 , 6-9 May 2001;

[5] *Klaue, J.; Gross, J.; Karl, H.; Wolisz, A.;* “Semantic-aware Link Layer Scheduling of MPEG-4 Video Streams in Wireless Systems”; Telecommunication Networks Group, Technical University of Berlin, Einsteinufer 25, 10587 Berlin, Germany, Proc. of Applications and Services in Wireless Networks (AWSN), Bern, Switzerland, July 2003;

[6] *Zhi-Li Zhang; Srihari Nelakuditi; Rahul Aggarwal, Rose P. Tsang;* “Efficient Selective Frame Discard Algorithms for Stored Video Delivery across Resource Constrained Networks”, INFOCOM 1999;

[7] *Daniel Fernàndez; Hèctor Montes;* “An Enhanced Quality of Service Method for Guaranteed Bitrate Services over Shared Channels in EGPRS System”; Vehicular Technology Conference, 2002. VTC 2002 Spring, May 2002;

[8] *Basso, A.; Byoung-Jo Kim; Zhimei Jiang;* “Performance evaluation of MPEG-4 video over realistic EDGE wireless networks”; Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on , Volume: 3 , 27-30 Oct. 2002;

[9] *Yuen, J.; Kam-Yiu Lam; Chan, E.;* “Fair and adaptive scheduling protocol for video stream transmission in mobile environment”; Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on , Volume: 1 , 26-29 Aug. 2002.