Multi-User Video Streaming over Multiple Heterogeneous Wireless Networks: A Distributed, Cross-Layer Design Paradigm

Liang Zhou, Benoit Geller, Xiaojun Wang, Anne Wei, Baoyu Zheng, Han-Chieh Chao

Multi-User Video Streaming over Multiple Heterogeneous Wireless Networks: A Distributed, Cross-Layer Design Paradigm

Liang Zhou, Benoît Geller, Xiaojun Wang, Anne Wei, Baoyu Zheng, and Han-Chieh Chao

Abstract

In this paper, we address the problem of QoS (Quality of Service) provisioning for multi-user video streaming over multiple heterogeneous wireless networks based on the distributed, cross-layer design framework. By jointly considering the rate allocation and the Joint Source-Channel Coding (JSCC), our proposal aims at maximizing the QoS provisioning under the given resource constraint. At first, we develop and evaluate a framework for optimal video rate allocation over multiple networks based on the observed Available Bit Rate (ABR) and the Round Trip Time (RTT) over each access network, as well as the video rate-distortion characteristics. The rate allocation is formulated as a convex optimization problem that minimizes the sum of all video streams expected distortion. Then, we propose an analytical JSCC scheme for error-resilient scalable encoded video, and integrate the JSCC with the specific rate allocation algorithm to improve the constructed video quality by optimally applying the appropriate channel coding rate given the constraints imposed by the transmission rate and the prevailing channel conditions. Objective and subjective simulation results are provided which demonstrate the effectiveness of our proposed joint scheme.

Index Terms

cross-layer design; QoS provisioning; video transmission; heterogeneous networks

L. Zhou and B. Zheng are with the Electronic Engineering Department, Shanghai Jiao Tong University, Shanghai, China and Institute of Signal Processing and Transmission, Nanjing University of Post and Telecommunications, Nanjing, China; B. Geller is with the LEI, ENSTA-ParisTech, Paris, France; X. Wang is with the School of Electronic Engineering, Dublin City University, Dublin, Ireland; A. Wei is with the LATTIS, University of Toulouse II, Toulouse, France; H.-C. Chao is with the department of Electronic Engineering and Institute of Computer Science and Information Engineering, National Ilan University, Taiwan, ROC.

ACRONYM AND NOTATION

ABR	Available Bit Rate
AIMD	Additive-Increase-Multiplicative-Decrease
CBR	Constant Bit Rate
FEC	Forward Error Correction
GOP	Group Of Picture
IMS	IP Multimedia Subsystems
JSCC	Joint Source-Channel Coding
MDC	Multiple Description Coding
NDRA	Novel Distributed Rate Allocation
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
RB	Residual Bandwidth
RTT	Round Trip Time
SPIHT	Set Partitioning In Hierarchical Trees
D	distortion, D_{all} is the overall distortion, D_{comp} is the distortion caused by source compression and D_{loss} is the distortion caused by packet loss
F	frame rate
$G(\sigma)$	distribution function giving the probability of the gap length greater than $\sigma - 1$
$g(\sigma)$	density function giving the probability of a gap length σ
I_s^n	number of coding layer for the user s over network n
K_s^n	source coding for the user s over network n in one GOP
L	packet size
L_B	average burst length
$\mathbf{N}$	set of heterogeneous wireless networks $\mathbf{N} = \{1, 2, ..., N\}$, which contains the number of N networks
N_{GOP}	number of the frames in one GOP
N_s^n	number of packets transmitted by user s over network n in one GOP
P_{loss}	total packet loss rate including the random packet loss and the packet loss caused by late arrival
P_B	average random packet loss rate
P_{BG},P_{GB}	denotes the probability from the state B to G and state G to B respectively in the Gilbert model
$P(n, m)$	probability of m errors with in a block of n symbols
$\mathbf{R}$	matrix of the allocated rate $\mathbf{R} = \{R_s^n\}_{S\times N}$, in which each element R_s^n corresponds to the allocated rate of user $s \in \mathbf{S}$ over network $n \in \mathbf{N}$
$R(n, m)$	probability of $m - 1$ erroneous symbols with the $n - 1$ symbols following an erroneous symbol.
$\mathbf{S}$	set of users $\mathbf{S} = \{1, 2, ..., S\}$, which contains the number of S users
T	delay constraint
$\mathbf{U}$	set of utility function $\mathbf{U} = \{U_s^n\}_{S\times N}$, in which each element U_s^n corresponds to the utility value of user $s \in \mathbf{S}$ over network $n \in \mathbf{N}$

I. Introduction

As multimedia is expected to be the major traffic source on the next-generation wireless networks, QoS (Quality of Service) provisioning for wireless video transmission has become a critically important issue. In addition, recent years have also witnessed the increasing efforts towards standardization of architectures for convergence of heterogeneous access networks, and moreover, the integration of heterogeneous networks has fully become part of the 4G network design [1]. IEEE 802.21 [2] is delineating a framework to enable handovers and interoperability between heterogeneous wireless and wireline networks. Therefore, supporting multimedia applications over heterogeneous networks has been one of the main fields of research in the networking and video coding communities. For example, the IMS (IP Multimedia Subsystems) platform [3] has defined an overlay architecture for providing multimedia services on top of heterogeneous wireless networks. Note that, the problem of video streaming over heterogeneous networks is further complicated by the heterogeneity of both the video contents and the network conditions. Up to now, providing a satisfactory communication quality in a heterogeneous wireless system is still a challenging problem because the end-to-end QoS guarantee is difficult to be provided [19].

The issue of supporting error-resilient video transport over error-prone wireless networks has received considerable attention recently. [4], [6], [10], and [20] presented some source coding-based approaches that divide the original bit-stream into multiple streams, called Multiple Description Coding (MDC), to tradeoff the error-resilience and the coding complexity; [13] and [9] amplified the benefits of using MDC by combining it with path diversity; in this context, each stream is explicitly transmitted over an independent path to the receiver in order to achieve higher tolerance to packet loss and delay due to network congestion. In [5], [14], and [23], the effect of different FEC (Forward Error Correction) coding schemes on reconstructed video quality had been investigated. In order to trade-off between the sustained quality of video stream and the network capacity, [8], [21], and [22] investigated the impact of the operating rate on the overall video quality; [15] discussed both centralized and distributed solutions for joint routing and rate allocation of multiple video streams in wireless ad hoc networks. Moreover, the rate adaptation of multimedia streams was studied in the context of heterogeneous networks in [7], where the authors proposed an architecture to allow online measurement of network characteristics and video rate adaptation via transcoding.

Typically, for real-time video communications over wireless networks, there are two main factors which can greatly affect the perceived video quality: the ***operating rate*** and the ***transmission error***. On one hand, for video streaming, high bandwidth requirements are coupled with tight delay constraints as packets need to be delivered in a timely fashion to guarantee continuous media playout. More specifically, if the operating rate is higher than the optimal transmission rate along a path, many packets will be discarded due to late arrival caused by congestion. On the contrary, if the operating rate is lower, performance loss will occur due to the source coding inefficiency. Hence, a rate control scheme is both desirable and necessary to achieve a satisfactory level of received video quality over wireless networks. On the other hand, transmission errors are generally caused by multi-path channel fading, interference from other electronic devices, and node mobility [23]. In addition, most of the video compression coding standards, including MPEG-4 and H.264, are designed to achieve high compression efficiency at the expense of error-resilience. This poses a severe problem, namely error propagation, where errors due to packet loss in a reference frame propagate to all of the dependent frames leading to visual artifacts that can be long lasting and annoying [16], [25]. To provide a reasonable QoS, it is important that the source coders be both error-resilient and network-adaptive. In order to achieve improved video quality supported by heterogeneous wireless networks, and to provide an overall more robust video delivery system, these two factors are jointly considered in this paper.

In this work, we explore the potential synergies of exchanging information between different layers to support video streaming over heterogeneous wireless networks. The main contributions and novelties of this paper are: (1) developing a framework for optimal video rate allocation over heterogeneous networks, based on the observed ABR (Available Bit Rate) and the RTT (Round Trip Time) over each network as well as the video rate-distortion characteristics; (2) proposing an analytical JSCC (Joint Source-Channel Coding) scheme for error-resilient scalable encoded video, in which the video sequence is encoded into multiple independent streams based on 3-D SPIHT (Set Partitioning In Hierarchical Trees) algorithm and each stream is assigned a FEC (Forward Error Correction) code to avoid error propagation; (3) integrating the JSCC with the specific rate allocation algorithm, which optimally applies the appropriate channel coding rate given the constraints imposed by the transmission rate obtained from the proposed rate

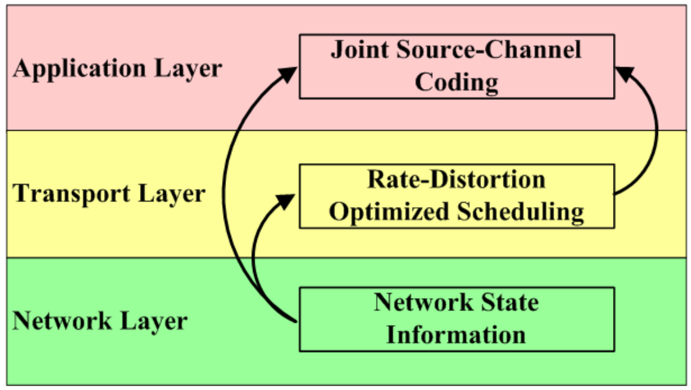


Fig. 1. Illustration of the cross-layer design of the total system.

allocation scheme and the prevailing channel condition. The combination of the rate allocation and the JSCC represents a cross-layer architecture as shown in Fig.1.

The rest of the paper is organized as follows. In Section II, we propose a cross-layer distributed rate allocation scheme for multiple video streaming sessions sharing multiple heterogeneous networks. In Section III, an analytically optimized JSCC is proposed given the optimal transmission rate and the prevailing channel conditions. We present some selected simulation results for the proposed joint scheme transmission over heterogeneous wireless networks in Section IV, and finally make some concluding remarks in Section V.

II. Cross-Layer Rate Allocation Scheme

In this section, we address the problem of rate allocation among multiple streams over multiple heterogeneous networks. At first, we propose a distortion model which captures both the impact of the encoder quantization and the packet loss on the overall video quality caused by the operation rate. Then, a distributed rate allocation scheme is presented to minimize the distortion, in which some cross-layer information exchange ensures that the allocated rates are updated according to changes in network conditions.

A. Distortion Model

In general, the reconstructed video quality is affected by both the source compression and the quality degradation due to packet losses caused by either transmission errors or late arrivals.

Here, we assume that the two forms of induced distortion are independent and additive. Thus, we can calculate the overall distortion D_{all} as:

$$D_{all} = D_{comp} + D_{loss} \tag{1}$$

where the distortion introduced by quantization is denoted by D_{comp}, and the additional distortion caused by packet loss is denoted by D_{loss}. According to [12], the distortion caused by source compression can be approximated by:

$$D_{comp} = \frac{\theta}{R - R^0} + D^0 \tag{2}$$

where R is the rate of the video stream, θ, R^0 and D^0 are the parameters of the distortion model which depend on the encoded video sequence as well as on the encoding structure. Using nonlinear regression technique, these parameters can be estimated from empirical rate-distortion curves obtained by encoding a sequence at different rates [11]. Likewise, the distortion caused by packet loss can be modeled by a linear model related to the packet loss rate P_{loss}:

$$D_{loss} = \alpha P_{loss} \tag{3}$$

where α depends on parameters related to the compressed video sequence, such as the proportion of intra-coded macro-blocks and the effectiveness of error concealment at the decoder [12]. The packet loss rate P_{loss} reflects the combined rate of random losses and late arrivals of video packets. In a bandwidth-limited network, this combined loss rate can be further modeled based on the M/M/1 queuing model. In this case, the delay distribution of packets over a single link is exponential [11]. Note that, since the end-to-end delay of packet delivery in wireless network is dominated by the queuing delay at the bottleneck link, the empirical delay distribution for realistic traffic patterns can still be modeled by an exponential formulation:

$$Pr\{Delay > T\} = e^{-\lambda T} \tag{4}$$

where $Pr\{\cdot\}$ denotes probability, T reflects the delay constraint and λ is the arriving rate which is determined by the average delay:

$$\lambda = \frac{1}{E\{Delay\}} \tag{5}$$

$E\{\cdot\}$ represents the expectation value. Generally, λ needs to be determined empirically from end-to-end delay statistics over the network. In order to present a general solution for online operation, here we construct a model to approximate the average packet delay.

Consider multiple wireless networks $\mathbf{N} = \{1, 2, ..., N\}$ simultaneously available to multiple users $\mathbf{S} = \{1, 2, ..., S\}$. Each network $n \in \mathbf{N}$ is characterized by its Available Bit Rate ABR^n and Round Trip Time RTT^n, which are measured and updated periodically. It should be noted that as channel conditions in wireless environments change on very short time scales (e.g., up to a few tens of ms), we assume that ABR^n and RTT^n represent average values computed on a larger time scale (e.g., one to a few seconds), and represent the average channel conditions for user $s \in \mathbf{S}$ on the given period.

Therefore, the rate allocation can be expressed in matrix form: $\mathbf{R} = \{R_s^n\}_{S\times N}$, where each element R_s^n corresponds to the allocated rate of user $s \in \mathbf{S}$ over network $n \in \mathbf{N}$. Consequently, the total allocated rate over network n is $R^n = \sum\limits_{s\in\mathbf{S}} R_s^n$, and the total allocated rate for user s is $R_s = \sum\limits_{n\in\mathbf{N}} R_s^n$. We denote RB^n, the Residual Bandwidth (RB) over network n, as:

$$RB^n = ABR^n - \sum_{s\in\mathbf{S}} R_s^n \tag{6}$$

From the perspective of user s in network n, the observed available bandwidth ABR_s^n is :

$$ABR_s^n = ABR^n - \sum_{s'\neq s, s'\in\mathbf{S}} R_{s'}^n \tag{7}$$

As the allocated rate on each network approaches the maximum achievable rate, average packet delay typically increases due to network congestion. We use a simple fractional function to approximate the non-linear increase of packet delay with traffic rate over network $n \in \mathbf{N}$, as:

$$E\{Delay\} = \frac{\beta^n}{RB^n} = \frac{\beta^n}{ABR^n - \sum\limits_{s\in\mathbf{S}} R_s^n} \tag{8}$$

which is reminiscent of the classical M/M/1 queuing model [24]. Assuming equal delay on both directions, the value of β^n can be estimated from the most recent observations of RTT'^n and RB'^n:

$$\beta^n = \frac{RB'^n \cdot RTT'^n}{2} \tag{9}$$

More specifically, if current residual bandwidth is equal to the past observation value for network $n \in \mathbf{N}$ ($RB'^n = RB^n$), the average current delay is $RTT'^n/2$. Therefore, for each network $n \in \mathbf{N}$

$$Pr\{Delay > T\} = e^{-\lambda T} = e^{-\frac{2(ABR^n - \sum\limits_{s\in\mathbf{S}} R_s^n)}{RB'^n RTT'^n} T} \tag{10}$$

Taking into account P_B^n, the average random packet loss rate in network $n \in \mathbf{N}$ due to transmission errors, the total packet loss rate in network $n \in \mathbf{N}$ is then:

$$P_{loss}^n = P_B^n + (1 - P_B^n)Pr\{Delay > T\} = P_B^n + (1 - P_B^n)e^{-\frac{2(ABR^n - \sum\limits_{s \in \mathbf{S}} R_s^n)}{RB'^n RTT'^n}T} \tag{11}$$

Therefore, the distortion from packet loss in network $n \in \mathbf{N}$ can thus be expressed as:

$$D_{loss}^n = \alpha P_{loss}^n = \alpha \left(P_B^n + (1 - P_B^n)e^{-\frac{2(ABR^n - \sum\limits_{s \in \mathbf{S}} R_s^n)}{RB'^n RTT'^n}T} \right) \tag{12}$$

B. Rate Allocation Algorithm

Based on the previous discussion, we seek to minimize the sum of the total distortion D_{all} as follows:

$$\min_{s \in \mathbf{S}, n \in \mathbf{N}} \quad \left\{ D_{all}(R_s^n) = \sum_{s \in \mathbf{S}} \left(\frac{\theta_s}{\sum\limits_{n \in \mathbf{N}} R_s^n - R_s^0} + D_s^0 \right) + \sum_{n \in \mathbf{N}} \alpha \left(P_B^n + (1 - P_B^n)e^{-\frac{2(ABR^n - \sum\limits_{s \in \mathbf{S}} R_s^n)}{RB'^n RTT'^n}T} \right) \right\} \tag{13}$$

$$\begin{aligned} subject\ to \quad & R_s^n = \frac{ABR_s^n}{\sum\limits_{n \in \mathbf{N}} ABR_s^n} R_s, \forall n \in \mathbf{N} \\ & R_s^n \leq ABR_s^n, \forall n \in \mathbf{N} \end{aligned}$$

where θ_s, R_s^0 and D_s^0 are the corresponding parameters for user $s \in \mathbf{S}$. Intuitively, reconstructed video quality is limited by coarse quantization at lower rates, whereas at high rates, the video stream will cause more network congestion. This, in turn, leads to higher loss rates and reduces the video quality. For video steaming in bandwidth-limited environments, we therefore expect to achieve maximum decoded quality for some intermediate rate.

In order to get the optimal result with fast convergence adapting to the online operation, we now propose a heuristic approach for solving the rate allocation optimization problem based on the utility framework introduced in [17], which iteratively takes a locally optimal decision on each user at each network. We define $\overline{R_s^n} \rightarrow R_s^n$ as the transition of the next allocation rate for the user $s \in$ **S** in network $n \in \mathbf{N}$, $\overline{R_s^n} = R_s^n + \triangle R_s^n$, where $\triangle R_s^n$ is the rate improvement varied at each iteration[1]. The utility of this transition can be computed as:

$$U_s^n = \frac{D_{all}(\overline{R_s^n}) - D_{all}(R_s^n)}{\overline{R_s^n} - R_s^n} \tag{14}$$

[1]In theory, the original $\triangle R_s^n$ can be chosen at random as long as it is less than ABR_s^n; here the original value of $\triangle R_s^n$ is set to $ABR_s^n/2$.

TABLE I

THE PROPOSED NDRA ALGORITHM

Input:
P_B^n, ABR_s^n, RTT^n $\forall$ user $s \in$ **S** in network $n \in$ **N**;
$R_s^n = 0$, $\triangle$ $R_s^n = ABR_s^n/2$, $\forall$ user $s \in$ **S** in network $n \in$ **N**;
Output:
Global Rate Allocation **R**
Procedure RateAllocation
while (true)
for s=1 to S **do**
    for n=1 to N **do**
        compute the utility of $\overline{R_s^n} \rightarrow R_s^n$:
        $U_s^n = \frac{D_{all}(\overline{R_s^n}) - D_{all}(R_s^n)}{\overline{R_s^n} - R_s^n}$;
        $\triangle$ R_s^n =$\triangle$ R_s^n / U_s^n;
        $\overline{R_s^n} = R_s^n + \triangle$ R_s^n;
    end for
end for
find $\mathbf{U}^* = \arg\max_{\mathbf{R}} \mathbf{U}$;
IntraNet(**R**, $\mathbf{U}^*$, n)
Procedure IntraNet(**R**, $\mathbf{U}^*$, n)
if network n has enough free resources **then**
    $\overline{R_s^n} \rightarrow R_s^n$;
    update free resources on network n;
else
    InterNet(**R**, $\mathbf{U}^*$, n);
end if
Procedure InterNet(**R**, $\mathbf{U}^*$, n)
find other user that can transfer part of its allocated rate to network $n' \neq n$ with maximum transition utility improvement $\triangle$ **U**;
if $\triangle$ **U** > 0 **then**
    perform the resource free up:
    $\overline{R_s^n} \rightarrow R_s^n$;
    update free resources on network n and n';
else
    break;
end if

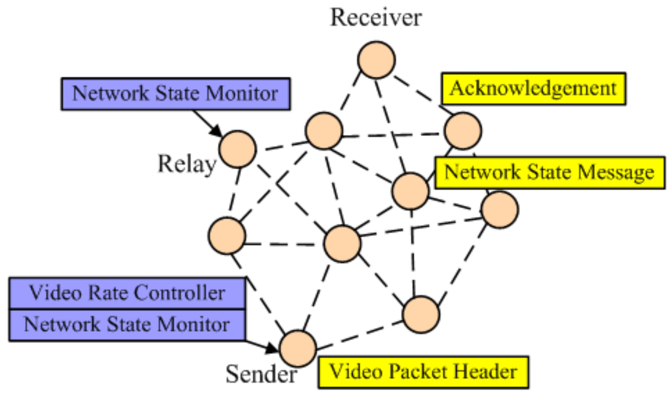


Fig. 2. Cross-layer information exchange between the network state monitor at the network layer and the video rate controller at the transport layer.

The total utility matrix is $\mathbf{U} = \{U_s^n\}_{S\times N}$. During each iteration, the proposed algorithm finds the $\mathbf{R} = \{R_s^n\}_{S\times N}$ that brings the highest utility $\mathbf{U}^* = \{U_s^n\}_{S\times N}$ to the overall system by its transition:

$$\mathbf{U}^* = \arg\max_{\mathbf{R}} \mathbf{U} \tag{15}$$

The proposed rate allocation algorithm is executed by each node (i.e. each user) and starts to allocate resources to user s in network n. Once the resources of the network n are depleted, the algorithm will find a different user that can free the required resources for user s in the other network, by allocating part of its rate. This operation is performed as long as the overall utility of the system is still improved, and as long as free network resources still exist in the overall system. The algorithm stops when there are no more free resources in the network system, or when no other possible user transition can bring any improvement in the overall system utility.

The proposed Novel Distributed Rate Allocation (NDRA) algorithm (see Table I) represents a sketch of the proposed algorithm. In this algorithm, the IntraNet procedure always attempts to increase the system's utility by allocating the resource in the network $n \in \mathbf{N}$ to the best user. If the free resources are not sufficient, the InterNet procedure tries to find a new user that can free up enough resources by allocating parts of its allocated rate through other network $n' \neq n \in \mathbf{N}$. As long as the network resources allow it, the whole procedure is repeated until no extra utility improvement can be brought to the overall system.

In order to adapt video source rates at the transport layer according to network states reported

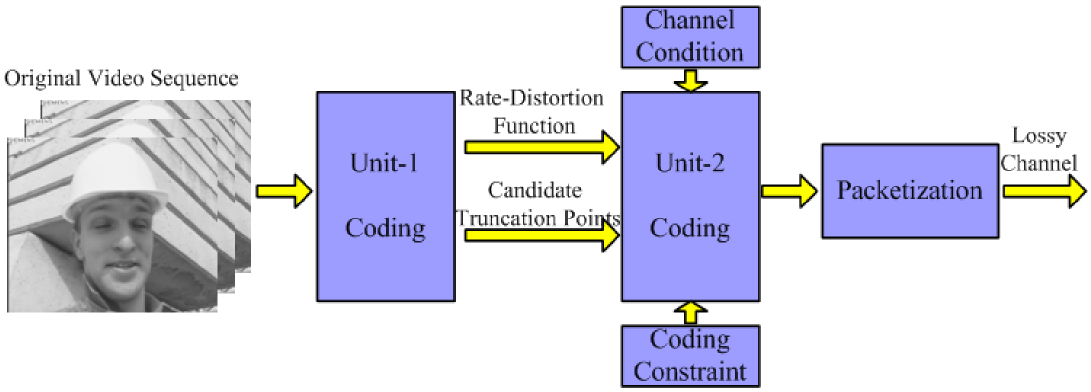


Fig. 3. Architecture of JSCC.

from the network layer, some cross-layer information exchange is needed. Fig.2 illustrates various components in such a system. At the network layer, the distributed allocation scheme would require track the observations of ABR^n and RTT^n over all available access networks. It also records the intended rate allocation R_s^n advertised by each user, and calculates the value of D_{all} and U_s^n accordingly. At the transport layer, the video rate controller at the source advertises its intended rate allocation R_s^n. The network state monitor traversed by the stream then calculates the relevant parameters based on its local cache of ABR^n, RTT^n and RB^n within the same access network. The destination node extracts such information from the video packet header and reports back to the sender in acknowledgment packets, so that the video rate controller can re-optimize its intended rate R_s^n based on the proposed NDRA algorithm, with updated network state information.

III. Proposed Joint Source-Channel Coding

In this section, we describe the application of JSCC approach subject to the allocated transmission rate obtained from the proposed rate allocation scheme. In what follows, we first introduce the architecture of JSCC, and describe the employed packet-loss pattern approximation to represent the channel packet-loss process. Then, we propose a JSCC algorithm to optimize the perceived video quality by optimally setting the channel coding rate.

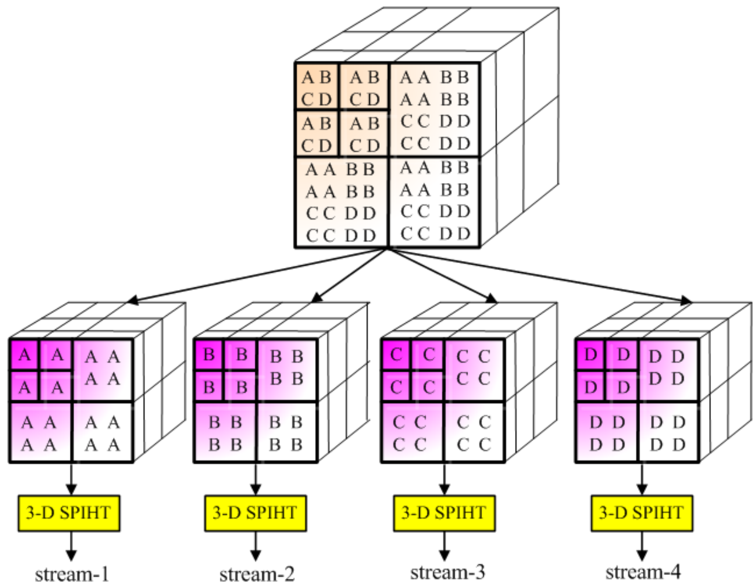


Fig. 4. Separating the 3-D wavelet transform coefficients into four streams.

A. The Architecture of JSCC

The proposed JSCC architecture contains two parts as shown in Fig.3. Unit-1 generates independent embedded streams using the 3-D SPIHT codec, while Unit-2 uses the coding constraints and channel conditions to pack the bit-streams into pack-streams of quality layers. This two-unit structure collects incremental contributions from the various streams into SNR (Signal-to-Noise Ratio) scalable quality layers in a way similar to that of embedded block coding with optimized truncation. The streams and rate-distortion functions generated by Unit-1 can be processed independently to channel conditions. The source and channel allocation algorithm in Unit-2 must be efficient to cope with the time varying channel conditions. In Unit-1, a video is divided into several independently encoded sections for additional functionality. The video coder divides the 3-D wavelet coefficients into multiple blocks according to their spatial and temporal relationships, and then encodes each block into sub-stream independently using the 3-D SPIHT algorithm.

Fig.4 shows an example of separating the 3-D wavelet transform coefficients into four independent blocks, each of which retains the spatiotemporal structure of 3-D SPIHT. Fig.5 shows how the concept of video layer is used in the encoding-decoding procedure. In this figure, dark

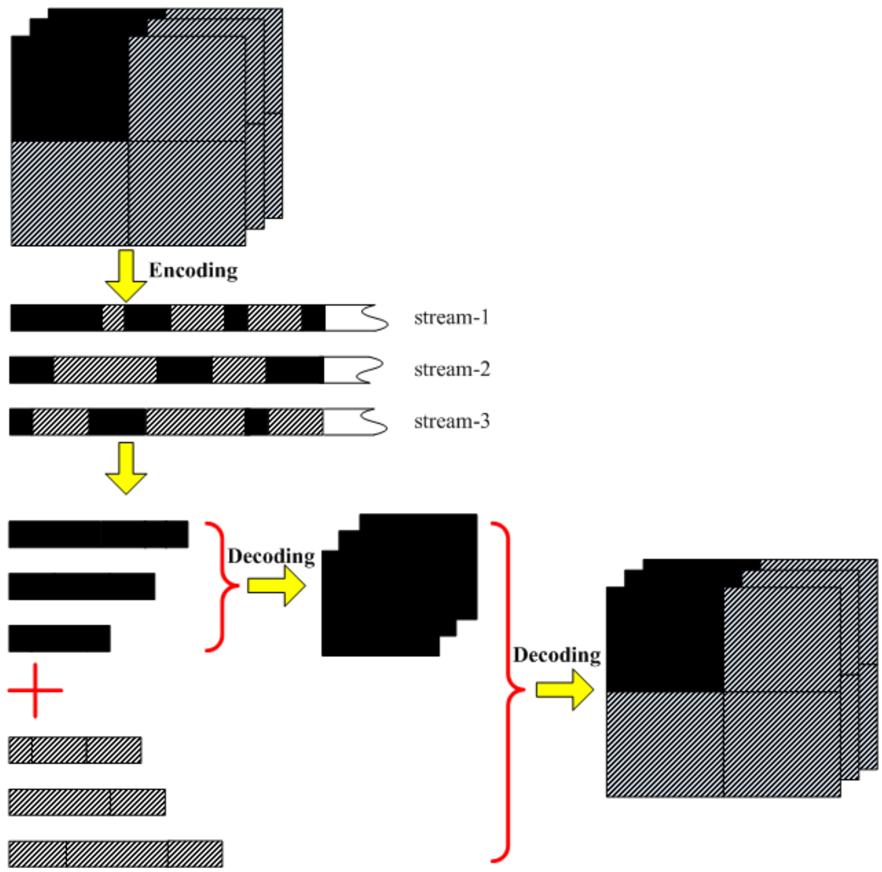


Fig. 5. The conception of video layer is used in the encoding-decoding procedure.

areas represent the low resolution of the video sequence, and the other parts are used for high resolution. After encoding some points of the video sequence, most of the remaining bit budget is used for encoding the higher frequency bands which contain video details not usually visible at reduced spatial-temporal resolution. Fig.5 illustrates this idea for a two-layer case of spatially scaled 352×288 video sequence. A low-resolution video can be decoded from the first layer only, and the frame size is 176×144. If we use three layers, the lowest layer's size would be 88×72.

B. Packet-Loss Pattern Approximation

For the RS (Reed-Solomon) correction code operating on b-bit symbols, the maximum block length is $2^b - 1$ symbols. A systematic (n, k) RS code appends $n - k$ redundant symbols to k

source symbols to make a block length n. The source symbols can be recovered correctly when the number of loss symbols is less than the minimum distance $d_{min} = n - k + 1$ of the code. The performance of a RS decoder $P_c(n, k)$ can be characterized by the code correct probability

$$P_c(n,k) = \sum_{m=0}^{n-k} P(n,m) \tag{16}$$

where $P(n, m)$ is the probability of m erasures within a block of n symbols. In a binary symmetric memoryless channel, we have

$$P(n,m) = \begin{pmatrix} n \\ m \end{pmatrix} P_B^m (1 - P_B)^{n-m} \tag{17}$$

where P_B is the average packet loss rate. For channels with memory, it is more complicated to calculate $P(n, m)$. Here, we use a two-state Markov model (i.e.Gilbert model) to simulate the bursty packet loss behavior. The two states of this model are denoted as G (Good) and B (Bad). In state G, packets are received correctly and timely, whereas, in state B, packets are assumed to be lost. This model can be described by the transition probabilities P_{GB} from state G to B and P_{BG} from state B to G. Then the average error probability P_B is given by

$$P_B = \frac{P_{GB}}{P_{GB} + P_{BG}} \tag{18}$$

and the average burst length

$$L_B = \frac{1}{P_{BG}} \tag{19}$$

which is the average number of consecutive symbol errors. The Markov model is a renewal model, and such models are determined by the distribution of error-free intervals, known as gap. Let the gap of length σ be the event that after a lost packet, $\sigma - 1$ packets are received correctly and then a packet is lost again. The gap density function $g(\sigma)$ gives the probability of a gap length σ. The gap distribution function $G(\sigma)$ is the probability of a gap length greater than $\sigma - 1$. These functions can be derived as [23]

$$g(\sigma) = \begin{cases} 1 - P_{BG}, & \sigma = 1 \\ P_{BG}(1 - P_{GB})^{\sigma-2} P_{GB}, & \sigma > 1 \end{cases} \tag{20}$$

$$G(\sigma) = \begin{cases} 1 - P_{BG}, & \sigma = 1 \\ P_{BG}(1 - P_{GB})^{\sigma-2}, & \sigma > 1 \end{cases} \tag{21}$$

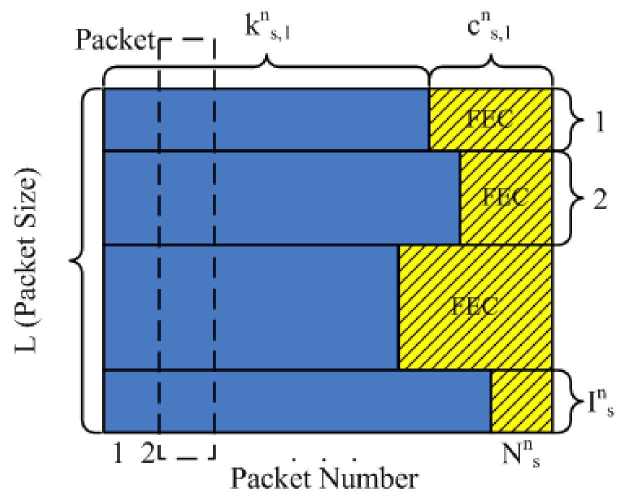


Fig. 6. Packetization of source-channel bits of different layers for user s in network n.

Let $R(n,m)$ be the probability of $m-1$ erroneous symbols within the next $n-1$ symbols following an erroneous symbol. It can be evaluated using the following

$$R(n,m)=\begin{cases} G(n), & m=1 \\ \sum_{\sigma=1}^{n-m+1} g(\sigma)R(n-\sigma,m-1), & 2\leq m\leq n \end{cases} \tag{22}$$

Then the probability of m errors within a block of n symbols becomes:

$$P(n,m)=\begin{cases} \sum_{\sigma=1}^{n-m+1} P_B G(\sigma)R(n-\sigma+1,m), & 1\leq m\leq n \\ 1-\sum_{\sigma=1}^{n} P(n,m), & m=0 \end{cases} \tag{23}$$

C. Proposed JSCC

Fig.6 shows how multiple encoded sequences of different quality layers are protected based on systematic RS codes. For notational convenience, we define layer 1 as the highest layer and layer I_s^n as the lowest layer to be sent for user $s\in\mathbf{S}$ in network $n\in\mathbf{N}$. Let N_s^n be the number of packets that are used to send the combined source data and redundancy for user $s\in\mathbf{S}$ over network $n\in\mathbf{N}$ in a GOP (Group Of Picture) and let L be the packet size in bytes. In this scheme, for user s in network n, the source bits belonging to layer i ($1\leq i\leq I_s^n$) are $k_{s,i}^n$ packets and the remaining $c_{s,i}^n=N_s^n-k_{s,i}^n$ packets are filled with channel coding redundancy. In other words, the source data for layer i is protected by RS$(N_s^n,k_{s,i}^n)$ code .

Our proposed coding scheme is performed on GOP basis. We define the total number of packets at period t to be sent from source $s \in \mathbf{S}$ for a GOP period $N_s(t)$ as

$$N_s(t) = \lceil \frac{R_s(t) \times N_{GOP}}{F \times L} \rceil \tag{24}$$

where $R_s(t)$ is the total rate in bytes/s at period t for the combination of data and redundancy for user s over all networks; N_{GOP} is the number of frames in one GOP; F is the frame rate in frames/s. In this work, we assume that each user transmits his stream over N networks ($N \geq 1$). Then the proposed algorithm divides $N_s(t)$ into $N_s^1(t)$, $N_s^2(t)$,..., $N_s^N(t)$ so as to maximize the expected quality at the receiver. $N_s^n(t)$ ($n \in [1, N]$) represents the number of packets transmitted by user s over network n for at GOP period t. Taking into account the rate of user s over network n, $N_s^n(t)$ should satisfy the following condition:

$$N_s^n(t) = \lceil \frac{R_s^n(t) \times N_{GOP}}{F \times L} \rceil \tag{25}$$

where $R_s^n(t)$ is the rate of user s over network n at the GOP period t, which can be obtained from the proposed rate allocation scheme described in the previous section.

In order to get the optimal expected quality, for each user s, given $R_s^n(t)$, it is necessary to find N_s^n and $K_s^n = (k_{s,1}^n, k_{s,2}^n, ..., k_{s,I_s^n}^n)$ for $n \in [1, N]$ to minimize the overall distortion function (13). However, the computational complexity of this classic two-fold optimization problem is usually huge and it is hard to operate online. To get around the difficulty, we make use of the video layer concept to approximate and simplify the total distortion function as follows:

$$PSNR(t) = \sum_{n=1}^{N} \sum_{l=1}^{I_s^n} \Bigg(\sum_{j=N_s^n - k_{s,l}^n + 1}^{N_s^n} P(N_s^n(t), j) \sum_{i=l}^{I_s^n} PSNR_i(t) \Bigg) \tag{26}$$

$$\begin{aligned} subject\ to \quad & \sum_{n=1}^{N} N_s^n(t) = N_s(t), \\ & N_s^n(t) = \lceil \tfrac{R_s^n(t) \times N_{GOP}}{F \times L} \rceil \end{aligned}$$

where $PSNR_i$ denotes the ith video layer Peak Signal-to-Noise Ratio (PSNR), which represents the image distortion at the receiver relative to the original video sequence. Its definition is:

$$PSNR(dB) = 10 \log_{10}(\frac{255^2}{MSE}) \tag{27}$$

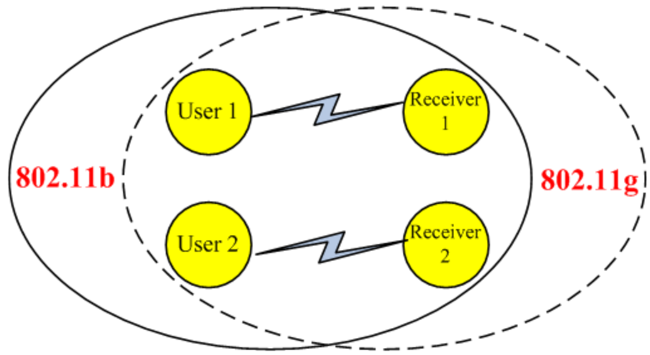


Fig. 7. Topology used for network simulations.

where MSE is the mean-square error between the original and the decoded luminance frame. $P(N_s^n(t), j)$ is the probability that j packets are lost out of $N_s^n(t)$ packets sent by user s over network n.

Each source and network independently runs the proposed rate allocation algorithm to get its optimal number of packets to transmit for a GOP period, using the information contained in the control packets that the receiver sends to all users. In theory, the proposed JSCC algorithm should try all possible combinations of (N_s^n, K_s^n) that satisfy the constraints in (26) and choose one that maximizes the expected quality. In order to get a fast solution, we set a predefined RS set which contains a number of RS codes acting as the potential channel codes. Of course, the computational complexity is proportional to the number of RS codes. According to our numerous experiments, the algorithm can achieve a good trade off between the performance and the computational complexity when N_s^n varies from 10 to 25 and K_s^n is more than half of the N_s^n. Therefore, for the finite number of networks, layers and the combination of (N_s^n, K_s^n), we can find a satisfying solution through exhaustive search.

IV. Simulation Results and Discussions

In this section, we conduct simulation experiments to study the performance of the proposed joint rate allocation and JSCC scheme in a distributed video streaming framework. First of all, we describe the simulation environment. Secondly, we present the main simulation results where we show the objective results of the performance of the proposed scheme under different scenarios. Finally, we conclude the section by summarizing the selected simulation results.

TABLE II

STATISTICS OF MEASURED ABR AND RTT

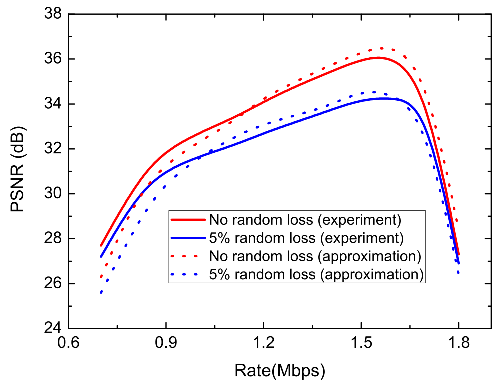
Network	Parameter	ABR (Mbps)	RTT (ms)
802.11b	Avg.	4.4	224.0
	Std. Dev.	1.3	8.7
802.11g	Avg.	15.8	297.0
	Std. Dev.	4.9	12.8

A. Simulation Environment

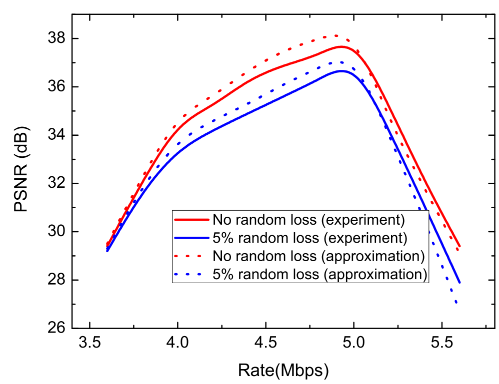
Here, we use the HD (High-Definition) *City* test sequence in our simulations to stream from two sources to two receivers with a maximum allowable total delay $T = 350$ milliseconds. The sequence has spatial resolution of 1280×720 pixels, and the frame rate is $F = 60$fps. The layer number is $I_s^n = 3$ and frame number in one GOP is $N_{GOP} = 30$ for all experiments. For simplicity, we use the Constant Bit Rate (CBR) model to represent video traffic. In this work, we simulate our joint scheme in NS-2 for an example network topology shown in Fig.7. Each user may stream video sequence via both two access networks. For simulation convenience, we don't consider the jitter parameter and just take into account the burst length L_B in this simulation. That is because this work focuses on the rate control combined with JSCC, and our simulation just aims at testing its efficiency.

Each network is simulated as a link with varying available bandwidth and delay, according to the traces collected from the actual access networks using the ABR and RTT measurements[2]. Table II summarizes the statistics of the collected ABR and RTT of each network trace over 200 GOP periods. During transmission, the environments are updated every frame transmission which can cause changes in the rate allocation and network resources; Within a frame transmission interval, the environment is kept constant. It should be noted that all the simulation results in this section have been obtained using 300 runs in order to get statistically meaningful average values.

[2]Forward and backward trip delays are both simulated as half of the measured RTT



(a) 802.11b



(b) 802.11g

Fig. 8. Decoded video quality approximate model and experimental data for HD *City* sequence at 60 frames per second and GOP length of 30 and a playout deadline of 350 ms. The value of α is 395 for both networks.

B. Performance Evaluation of the Proposed Scheme

At first, we validate the distortion model introduced in Subsection II.A. Fig.8 shows the rate-PSNR tradeoff when one user streams the HD video sequence *City* (300 frames) over the simulated two wireless networks which are noted as 802.11b (n=1) and 802.11g (n=2) network, respectively. The corresponding network parameters ABR (Mbps) and RTT (ms) for 802.11b and 802.11g are 4.5, 221 and 16.1, 290, respectively. The model is fit to experimental data for two cases: in the first case, the only losses considered are due to late arrivals; in the second, an additional end-to-end random loss rate of 5% is considered. The bell-shape of the curves illustrates that the highest performance is obtained when the streaming rate achieves the optimal tradeoff between compression quality and self-inflicted congestion. The approximate optimal operating rate computed by numerically solving (13) matches closely with experimental data.

To demonstrate the effectiveness of our proposed NDRA algorithm, we use the representative drop-tail scheme which employs the fixed rate allocation and the Additive-Increase-Multiplicative-Decrease (AIMD)-based rate allocation method which is used by TCP congestion control [18] for comparison. In order to get a clear picture of how the allocated rate reflects the reconstructed quality, we just use one user streaming over 802.11b network in this simulation. More specifically, the drop-tail scheme employs a fixed source coding rate $R_f = 1.50$Mbps and when the rate exceeds the current optimal transmission rate available for the selected source-destination pair, it will drop the subsequent encoded packets. The AIMD-based scheme probes the network for available bandwidth and reduces rate allocation after congestion occurs. Each user s initiates its rate at a specified rate R_s^{AIMD} corresponding to the minimum acceptable video quality, and increases its allocation by $\triangle\ R_s$ every $\triangle\ t$ seconds unless network congestion is perceived, in which case the allocated rate is dropped by $(R_s^n - R_s^{AIMD})/2$ over the congested network n. In the process of simulation, the increase in rate allocation is allocated to all available networks in proportion to the average of each ABR. In addition, congestion over network n is indicated upon detection of the lost packets, or when the observed RTT exceeds a specified threshold, based on the playout deadline of the video stream.

In Fig.9, we show a performance comparison between our proposed rate allocation scheme and the competing methods, drop-tail and AIMD, in the scenario where packet losses are caused only by channel over-pumping[3]. It should be noted that due to the use of CBR encoding, the video quality is not constant and varies periodically [14]. In Fig.9, the average PSNR using the proposed NDRA algorithm is 35.33 dB while it is 35.01 using AIMD-based method and 34.70 dB for the case of drop-tail. Thus, using the proposed NDRA algorithm can achieve almost 0.32 dB and 0.63 dB performance gains comparing to the AIMD and drop-tail scheme, respectively. From the network profile, illustrated in Table III (the value is averaged over one GOP), we can see that for GOP No.1, No.2, and No.4, the allocated transmission rate using the proposed NDRA method is higher than the fixed 1.50 Mbps. Thus, using rate allocation can fully exploit the reasonable transmission rate resulting in improved performance compared to using a fixed-rate coding scheme. On the other hand, for GOP No.3, it is obvious that the fixed source coding rate is higher than the allocated transmission rate; therefore, packet losses will occur when the transmission buffer is full resulting in the last couple of frames being lost which cause substantial performance degradation. A lost frame is concealed by just copying the previous frame and if several consecutive frames are lost, the degradation will be even more serious since the concealed frames are then used as correctly received frames to conceal the subsequent lost frames. This results in substantial error propagation. For example, in Fig.9, we can see that there is substantial performance degradation around the 90th frame for the no-rate

[3]Here, we assume that no transmission errors occurred

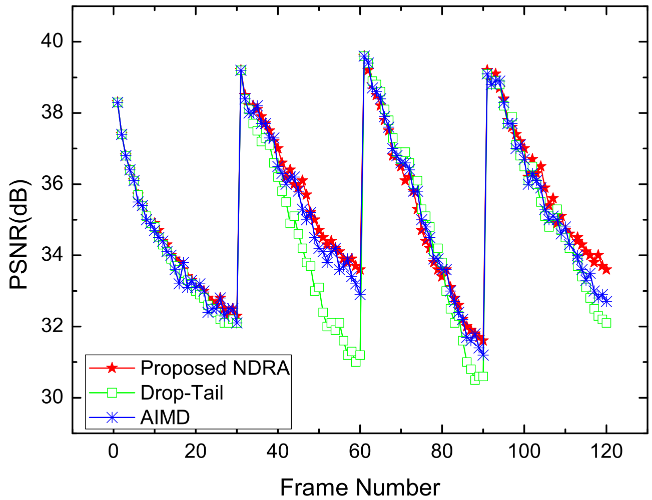


Fig. 9. Performance comparison between different rate allocation schemes.

control case due to channel over-pumping. Furthermore, although the performance degradation caused by the channel over-pumping packet losses has been partially compensated using passive error concealment, the performance is still not as good as using the rate allocation scheme. It should also be noted from Fig.9 that the performance achieved by the proposed NDRA method is also super to the traditional AIMD-based method. On one hand, although the AIMD-based method is adaptive to the change of the network conditions, the network is so dynamic that a congested node forwarding a few seconds might not be used at all when the source reacts to the congestion. On the other hand, the proposed NDRA method further take advantage of explicit knowledge of the video distortion-rate characteristics, and can achieve more balanced video quality.

Then, to evaluate the performance of the joint scheme, the proposed NDRA+JSCC scheme is benchmarked against other two competitive schemes: 1) NDRA+RS(15,9), in which a fixed RS (15, 9) code is used exclusively in the total system; 2) NDRA only, in which no channel coding is employed in the video transmission system. In Fig.10, we illustrate a plot of the PSNR versus the frame number for the test sequence; the corresponding network profile is illustrated in Table IV (the values are averaged over one GOP). The proposed JSCC scheme clearly achieves

TABLE III

CORRESPONDING CHANNEL PROFILE TO FIG.9(MBPS)

GOP No.	1	2	3	4
R_1^1(NDRA)	1.54	2.02	1.21	1.86
R_1^1(AIMD)	1.53	1.88	1.29	1.81
R_f	1.50	1.50	1.50	1.50

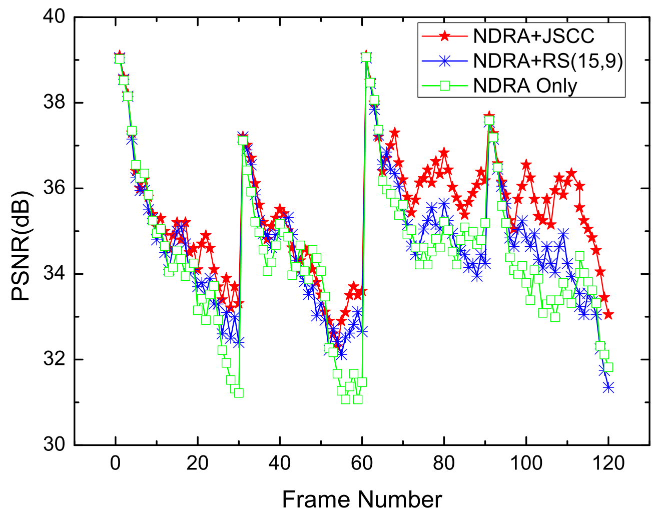


Fig. 10. Performance comparison between the proposed JSCC scheme with the representative fixed channel coding and without channel coding scheme.

TABLE IV

CORRESPONDING CHANNEL PROFILE TO FIG.10

GOP No.	1	2	3	4
R_1^1 (Mbps)	1.1	1.5	2.0	1.7
R_1^2 (Mbps)	3.8	5.3	4.6	4.4
R_2^1 (Mbps)	1.3	1.0	0.8	1.2
R_2^2 (Mbps)	5.5	4.1	5.1	5.3
P_B (%)	5.3	9.6	8.4	3.2
L_B	4.6	5.4	4.4	3.8

(a) (b) (c)

Fig. 11. Subjective comparisons of decoded frames for Susie sequence. (a) NDRA+JSCC; (b) NDRA+RS(15,9); (c) NDRA only (PSNR=33.7 dB, 31.3 dB, 29.7 dB, respectively).

a higher performance in terms of end-to-end PSNR compared to the no channel coding and the fixed RS (15, 9) schemes. From Fig.10, we can see that although using the RS (15, 9) code can provide protection, a substantial performance loss is observed compared to the proposed JSCC scheme which is so flexible that it can achieve an optimal value according to current condition. Also, although using no channel coding results in the best source coding efficiency, however, there are no error correcting capabilities which can be used to combat transmission errors; therefore, the corresponding performance is considerably worse than the JSCC scheme.

In order to provide a more comprehensive evaluation of the proposed joint scheme, we repeat the results for other video sequences of varying content complexity (see Table V) under the same simulation configuration as the previous experiments. From Table V, it can be observed that the proposed NDRA+JSCC scheme performances are superior to the other competitive methods, which is due to the network-adaptive and error-resilient characteristics of the proposed joint scheme. The previous objective results are based on a quantitative assessment of reconstructed PSNR values. In Fig. 11, we also display some subjective results based on the reconstructed frames taken from the decoded test sequence of the simulation run. From Fig. 11, we can see that the proposed NDRA+JSCC scheme can provide improved subjective performance compared to the other competing schemes.

C. Observations

Based on the selected objective and subjective simulation results described above, several main observations can be drawn:

TABLE V

PERFORMANCE COMPARISON FOR OTHER SEQUENCES UNDER DIFFERENT SIMULATION CONDITIONS

Video Sequence	T (ms)	PSNR of Different Methods (dB)				Improvement (dB)
		JSCC Only	NDRA Only	NDRA+RS(15,9)	NDRA+JSCC	
Harbor	300	33.2	31.9	33.6	34.9	1.3∼3.0
	400	33.5	31.9	34.1	35.7	1.6∼3.8
Crew	300	31.4	31.8	32.6	33.2	0.6∼1.8
	400	31.4	31.8	33.0	33.4	0.4∼2.0
Football	300	34.8	34.4	35.0	35.2	0.2∼0.8
	400	34.7	34.4	35.1	35.3	0.2∼0.9
Swim	300	33.9	34.3	34.3	35.6	1.3∼1.7
	400	33.9	34.6	35.1	35.9	0.8∼2.0
Ships	300	33.1	33.5	33.8	34.9	1.1∼1.8
	400	33.7	33.6	33.8	34.9	1.1∼1.3
Susie	300	29.4	30.8	32.2	33.1	0.9∼3.7
	400	29.4	30.4	32.2	33.4	1.2∼4.0

- The rate allocation plays an important role in the whole video transmission system. If the operating rate is lower than the optimal transmission rate, performance loss is due to the source coding inefficiency resulting from the use of an unnecessarily lower source coding rate; if the operating rate is higher, performance loss is caused by packet losses due to buffer overflow and network congestion.
- The JSCC actually achieves a higher performance. On one hand, the proposed JSCC approach is so flexible that it achieves an optimal PSNR value according to current condition; on the other hand, it achieves good trade-off between the performance and the source coding efficiency compared to the fixed channel coding schemes.
- The NDRA method combined with JSCC outperforms the competing methods from the objective comparisons.

V. CONCLUDING REMARKS

In this paper, we use distributed and cross-layer design to maximize the perceived video quality by combining rate allocation with joint source channel coding techniques. As detailed in the paper, our proposed joint scheme can adaptively respond to dynamic network conditions by joint adjusting the allocated rate in each network and the channel coding rate for each stream. The simulation results demonstrate the effectiveness of our proposed joint scheme for multiple video users over multiple heterogeneous wireless networks.

ACKNOWLEDGMENTS

This work was supported in part by the International Project PRA-SI (financed by France and China governments) under Grant No. SI04-03, and the Key Project of Nature Science Foundation of Jiangsu (China) under Grant BK2007729.

REFERENCES

[1] P. Vidales, J. Baliosion, J. Serrat, G. Mapp, F. Stejano, and A. Hopper, "Autonomic system for mobility support in 4G networks", IEEE Journal on Selected Areas in Communications, vol. 23, pp. 2288-2304, 2005.

[2] "IEEE 802.21", http://www.ieee802.org/21/.

[3] A. Cuevas, J. I. Moreno, P. Vidales, and H. Einsiedler, "The IMS platform: A solution for next generation network operators to be more than bit pipes", IEEE Communications Magazine, Special issue on Advances of Service Platform Technologies, vol. 44, pp. 75-81, 2006.

[4] Y. Wang, A.Reibman, and S.Lin,"Multiple description coding for video communications", Proc. IEEE, vol.93, pp.57-70, 2005.

[5] C.M. Fu, W.L. Hwang, and C.L. Huang, "A Joint Source and Channel Coding Algorithm for Error Resilient SPIHT-Coded Video Bitstreams", Journal of Visual Communication & Image Representation, vol.17, pp. 1164-1177, 2006.

[6] E. Akyol, A.M. Tekalp, and M.R. Civanlar , "A Flexible Multiple Description Coding Framework for Adaptive Peer-to-Peer Video Streaming", IEEE Journal of Selected Topics in Signal Processing, vol.1, no.2, pp.231-245, 2007.

[7] A. Szwabe, A. Schorr, F. J. Hauck, and A. J. Kassler, "Dynamic multimedia stream adaptation and rate control for heterogeneous networks", Journal of Zhejiang University, Science A, vol. 7, no. S1, pp. 63-69, 2006.

[8] E.Setton, B.Girod,"Rate-distortion analysis and streaming of SP and SI frames", IEEE Trans. Circuits Syst. Video Techn., vol.16, no.6, pp. 716-732, 2006.

[9] R. Tian, Q. Zhang, Z. Xiang, Y. Xiong, X. Li, W. Zhu, "Robust and efficient path diversity in application-layer multicast for video streaming", IEEE Transactions on Circuits and Systems for Video Technology, vol.15, no.8, pp.961-972, 2005.

[10] G. Wang, S. Futemma, E. Itakura, "Multiple-Description Coding for Overlay Network Streaming", IEEE Multimedia, vol.14, no.1, pp.74-82, 2007.

[11] X. Zhu, P. Agrawal, J. P. Singh, T. Alpcan, and B. Girod, "Rate Allocation for Multi-User Video Streaming over Heterogenous Access Networks", Proc. ACM Multimedia, pp. 37-46, Augsburg, Germany, September 2007.

[12] K. Stuhlmuller, N. Farber, M. Link and B. Girod, "Analysis of video transmission over lossy channels", IEEE Journal Select. Areas Commun., vol.18, no.6, pp. 1012-1032, 2000.

[13] J. Kim, R. M. Mersereau, Y. Altunbasak, "Distributed Video Streaming Using Multiple Description Coding and Unequal Error Protection", IEEE Transactions on Image Processing, vol.14, no.7, pp. 849-861, 2005.

[14] Q. Qu, Y. Pei, and J. W. Modestino, "An Adaptive Motion-Based Unequal Error Protection Approach for Real-Time Video Transport Over Wireless IP Networks", IEEE transactions on multimedia, vol. 8, no. 5, pp.1033-1044, 2006.

[15] X. Zhu, J.P. Singh, and B. Girod, “Joint Routing and Rate Alloction for Multiple Video Streams in Ad Hoc Wireless Networks”, Journal of Zhejiang University, Science A, vol. 7, no. 5, pp. 727-736, May 2006.

[16] M. G. Martini , M. Mazzotti, L. B. Catherine, J. Huusko, P. Amon, “Content adaptive network aware joint optimization of wireless video transmission”, IEEE Communications Magazine, vol.45, no.1, pp.84-90, 2007.

[17] D. Jurca, W. Kellerer, E. Steinbach, S. Kahn, S. Thakolsri and P. Frossard, “Joint Network and Rate Allocation for Video Streaming over Multiple Wireless Networks”, IEEE International Symposium on Multimedia, Taichung, Taiwan, December, 2007.

[18] E. Altman, K. Avrachenkov, C. Barakata, P. Dube, “Performance analysis of AIMD mechanisms over a multi-state Markovian path”, Computer Networks, vol.47, no.3, pp.307-326, 2005.

[19] I. F. Akyildiz, S. Mohanty, J. Xie, “A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems”, IEEE Communications Magazine, vol.43, no.6, pp.S29-S36, 2005.

[20] I. Radulovic and P. Frossard, “Multiple Description Coding with Redundant Expansions and Application to Image Communication” EURASIP Journal on Image and Video Processing, vol. 2007, Article ID 24863, 15 pages, 2007.

[21] J. Chakareski and P. Frossard, “Rate-Distortion Optimized Distributed Packet Scheduling of Multiple Video Streams over Shared Communication Resources”, IEEE Transactions on Multimedia, Special Issue on Distributed Media Technologies and Applications, vol. 8, no 2, pp. 207-218, April 2006.

[22] C. Vleeschouwer and P. Frossard, “Dependent packet transmission policies in rate-distortion optimized media scheduling”, IEEE Transactions on Multimedia, vol. 9, no 6, pp. 1241-1258, October 2007.

[23] S.-H. Chan, X. Zheng, Q. Zhang, W.-W. Zhu and Y.-Q. Zhang, “Video Loss Recovery With FEC and Stream Replication”, IEEE Transactions on Multimedia, vol. 8, No. 2, pp. 370-381, April 2006.

[24] S. C. Draper, M. D. Trott, G. W. Wornell, “A universal approach to queuing with distortion control”, IEEE Transactions on Automatic Control, vol.50, no.4, pp. 532-537, 2005.

[25] H.P. Shiang, M.Schaar, “Multi-user video streaming over multi-hop wireless networks: a distributed, cross-layer approach based on priority queuing”, IEEE Journal Select. Areas Commun., vol.25, no.4, pp. 770-785, May 2007.