Distributed Source Coding for WWAN Multiview Video Multicast with Cooperative Peer-to-peer Repair

Zhi Liu,   Gene Cheung,   Yusheng Ji

The Graduate University for Advanced Studies, National Institute of Informatics

2-1-2, Hitotsubashi, Chiyoda-ku

Tokyo, Japan 101-8430

Email: {liuzhi, cheung, kei}@nii.ac.jp

Abstract—**Video multicast over Wireless Wide Area Networks (WWAN) is difficult because of unavoidable packet losses and impracticality of retransmission on a per packet, per client basis, due to the known NAK implosion problem. Recent approach exploits clients' cooperation for packet recovery, so that a peer group's received WWAN packets are shared using a secondary network like Wireless Local Area Network (WLAN). For multiview video multicast, where a client can switch views interactively by subscribing to different WWAN multicast channels streaming different views, two new difficulties arise. First, system must provide timely view-switching mechanism, so that client can switch to a desired view quickly for correct decoding and display. Second, it is difficult for system to leverage neighboring peers for cooperative loss recovery, since neighbors are more likely to be subscribing to different views from a loss-stricken peer.**

In this paper, we use Distributed Source Coding (DSC), a new compression tool in video coding, to solve both problems. Each DSC frame is encoded with a set of predictor frames, and correct decoding only requires one of the predictors in the set to be available at decoder. Periodic insertion of DSC frames into video streams then enables a peer to switch from view v to v' at the DSC frame boundary, assuming DSC frame of view v' was encoded using a frame in view v as a predictor. For the same assumption, a neighbor watching view v can help a peer watching view v' evade error propagation resulting from earlier losses and resume decoding at the DSC boundary. Experiments show that optimized usage of DSC frames in a coding structure, where unequal error protection is enabled to decrease the probability of decoding failure earlier in a group of pictures, outperforms a structure using I-frames instead for view switching by up to 11dB in video quality in typical WWAN network loss environment.

I. INTRODUCTION

Video multicast over Wireless Wide Area Networks (WWAN) is challenging, given stringent playback deadlines of real-time video and unavoidable packet losses due to shadowing, channel fading and inter-symbol interference. Moreover, a streaming server cannot tailor retransmission for every lost packet experienced by each client due to the well-known NAK implosion problem [1]. To overcome channel losses, typical WWAN video multicast schemes [2] employ a large amount of Forward Error Correction (FEC) packets for channel coding, expending precious WWAN resource.

One recent approach to alleviate the wireless packet loss problem is *cooperative communication* [3], [4]. In short, given

the broadcast nature of wireless transmission (each transmission is heard by multiple receivers) and the "uncorrelatedness" of receivers' channels (hence unlikely for all receivers to undergo bad channel fades at the same time), peers that currently experience good channels (rich peers) can relay overheard information to clients who currently experience channel losses (poor peers), *without* relying on server retransmission. In the case of WWAN multicast, if a secondary network such as an ad hoc Wireless Local Area Network (WLAN) connects the peers locally [3], a WWAN multicast packet received by rich peers can be relayed to poor peers via WLAN without interrupting primary WWAN transmission. We term this WLAN local recovery process *cooperative peer-to-peer repair* (CPR).

An orthogonal development recently is multiview video technologies. Because of the continuing cost reduction of consumer-level cameras, a video sequence can now be recorded by a large array of cameras [5]; i.e., at each time instant, images of the same scene are simultaneously captured by multiple cameras from different viewpoints. Given encoded multiview content at the server, in an *interactive multiview video streaming* (IMVS) scenario [6], a viewer can interactively switch views by re-subscribing to different WWAN multicast channels streaming different time-synchronized views, so that only frames of interested viewpoint are received. While IMVS offers viewers a new interaction (view switching), it creates two implementation problems. First, system must provide timely view-switching mechanism, so that client can switch to a desired view quickly for decoding and display, without waiting for the next random-access I-frame in the targeted channel[1]. Second, it complicates the aforementioned cooperative packet recovery process; since viewers can now select views during a video multicast, a viewer may not have a neighboring peer watching the same view who can locally relay lost packets via CPR.

In this paper, we propose to use *distributed source coding* (DSC), a new compression tool in video coding, to solve

---

[1]Increasing the insertion frequency of intra-coded I-frames—necessary for new client joints—into a video stream can reduce view-switching delay (and evade error propagation), but their large sizes mean little WWAN bandwidth is left over for FEC packets, leaving the stream vulnerable to losses.

both problems. Each DSC frame is encoded using a set of predictor frames, and correct decoding only requires: i) correct reception of the DSC frame itself, and ii) correct decoding of *one* of the predictors in the set at the decoder buffer. Thus, periodic insertion of DSC frames into video streams enables a peer to switch from view v to v' at the next DSC frame boundary (which are more frequent than random-access I-frame boundaries), given DSC frame of view v' was encoded using a frame in view v as one of its predictors. Further, when a peer watching view v' experiences irrecoverable WWAN packet losses leading to error propagation in the stream (due to differential video coding), a neighbor watching view v can locally share its predictor frame via CPR to help the loss-stricken peer evade such error propagation and resume decoding at the next DSC boundary.

Moreover, because both the size and repair power of a DSC frame grows with the size of predictor set used during encoding, predictor sets for inserted DSC frames early in a group of pictures (GOP) can be selected to be larger than predictor sets for DSC frames later in the GOP. This results in unequal error protection (UEP) to further decrease the probability of failure earlier in the GOP. Experiments show that an optimized structure using DSC outperforms a structure using I-frames as switching mechanism by up to 11dB in video quality under typical WWAN network loss environment.

The outline of the paper is as follows. We first discuss related works in Section II. We then outline the multiview video multicast system and present our proposed coding structure using DSC in Section III. The problem of finding optimized parameters for our proposed DSC-based coding structure is formalized in Section IV, and the corresponding optimization is presented in Section V. Results and conclusion are presented in Section VI and VII, respectively.
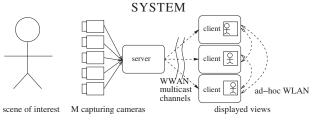
II. RELATED WORK

Because the probability of packet loss in WWAN can be substantial, conventional WWAN video multicast schemes [2] involve a large overhead of FEC packets. Our previous work on CPR [3] differs from these traditional approaches by utilizing a secondary ad hoc WLAN network for local recovery of packets lost in the primary network exploiting peers' cooperation. We have shown substantial gain in visual quality using CPR over non-cooperative schemes.

We stress that our assumption of available multi-homing capable devices (each has multiple network interfaces to connect to orthogonal delivery networks simultaneously) is a common one in the literature [7], [8], [9], where different optimizations are performed exploiting the multi-homing property. [7] shows that aggregation of an ad hoc group's WWAN bandwidths can speed up individual peer's infrequent but bursty content download like web access. [8] proposes an integrated cellular and ad hoc multicast architecture where the cellular base station delivered packets to proxy devices with good channel conditions, and then proxy devices utilized local ad hoc WLAN to relay packets to other devices. [9] shows that smart striping of FEC-protected delay-constrained media packets

across WWAN links can alleviate single-channel burst losses, while avoiding interleaving delay experienced in a typical single-channel FEC interleaver. Our current work extends this body of multi-homed literature by optimizing a different application: WWAN multiview video multicast.

DSC [10] has become a popular coding paradigm in the signal processing community. Exploiting DSC frame's unique decoding property (correctly decodeable if at least one in a set of predictor frames is correctly decoded in decoder buffer), previous works have found DSC to be useful in streaming applications such as unicast IMVS in wired networks [6]. To the best of our knowledge, this work is the first attempt to use DSC as mechanism for both view switching and loss recovery in a WWAN video multicast environment. Moreover, optimizing DSC for unequal error protection in a cooperative loss recovery setting is new.

III. WWAN MULTIVIEW VIDEO MULTICAST SYSTEM

Fig. 1. Overview of WWAN Multiview Video Multicast System

We first overview our WWAN multiview video multicast system. We then present our proposed coding structure using DSC for encoding of multiview video. Finally, we describe network assumptions for both WWAN (used for server-to-peer transmission) and ad hoc WLAN (used for cooperative peer-to-peer loss recovery).

A. System Overview

The components of our proposed WWAN multiview video multicast system, shown in Fig. 1, are the following. M cameras in a one-dimensional array captured a scene of interest from different viewing angles, and a server compresses video into a particular coding structure (to be discussed). Server transmits different video views, synchronized in time, in different WWAN multicast channels such as Multimedia Broadcast/Multicast Service (MBMS) in 3GPP [11]. Note that while multicasting views of video to a large group of clients is more bandwidth-efficient than unicasting one selected stream for each individual client, it is also more loss-prone due to lack of per-client retransmission, link adaptation [12], etc.

A peer interested in a particular view will subscribe to the corresponding multicast channel[2] and can switch to a neighboring view interactively by switching multicast channels every t/FPS seconds (an *epoch*), where FPS is playback speed of the video in frames per second. Fig. 1 shows three peers subscribing to three different multicast channels and resulting in three different views.

[2]We assume that due to power constraint, a peer will only subscribe to the one channel multicasting view of interest and decode frames of that one view.

Peers are connected to their neighbors via ad hoc WLAN, providing a secondary network for potential CPR frame recovery. If a neighbor to a peer is watching the same view v, then he can assist in frame recovery of same view v by relaying its own received packets via CPR. If neighbor is watching a different view v', then he cannot help recovery of the lost frames in view v. However, the neighbor can still help peer evade error propagation due to earlier frame losses by sharing its decoded frame in view v', so that peer's DSC frame in view v can be correctly decoded, assuming the DSC frame in view v was encoded using the received decoded frame of view v' as one of its predictors.

The WWAN server first multicasts one epoch worth of video to peers. Then during WWAN transmission of the next video epoch, cooperative peers will exchange received packets or decoded frames of the first video epoch. When the server multicasts the third video epoch, peers repair the second video epoch, and video in the first epoch is decoded and displayed. View-switching delay is hence two epochs $2t/FPS$. We assume the maximum tolerable view-switching delay τ seconds is pre-determined by the application, and by setting $t = \tau * FPS/2$, our system is one solution that satisfies this requirement.
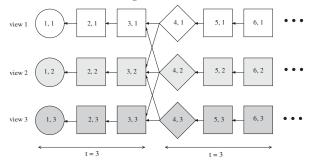
B. Multiview Video Coding Structure

Fig. 2. Example of coding structure for $M = 3$ views and coding unit of size $t = 3$. Circles, squares and diamonds are I-, P- and DSC frames, respectively. Each frame $F_{i,v}$ is labeled by its time index i and view v.

We assume the M available views are captured by cameras of close physical proximity so that strong spatial correlation exists among them. Let the number of source packets of frame $F_{i,v}$, time-indexed i of any view v, be s_i^I, s_i^P or s_i^{DSC} packets, respectively, when encoded as I-, P- or DSC frame. We propose to encode a GOP of $T = it$ frames, $i \in \mathcal{I}$, as follows: for each view v, encode a starting intra-coded I-frame $F_{1,v}$ with $t - 1$ trailing P-frames, each motion-compensated from previous frame, followed by one DSC frame $F_{t+1,v}$ and $t - 1$ trailing P-frames, following by another DSC frame $F_{2t+1,v}$ and $t - 1$ trailing P-frames, etc. A frame group composed of ith DSC frame $F_{it+1,v}$ and $t - 1$ trailing P-frames is termed a *coding unit* i of view v, $U_{i,v}$. Units $U_{0,v}$'s are special case with starting I-frames instead. See Fig. 2 for an illustration.

We encode each DSC frame $F_{it+1,v}$ using decoded P-frames $F_{it,u_i(v)}, \ldots, F_{it,w_i(v)}$ of previous time instant it and views $u_i(v), \ldots, w_i(v)$ as predictors[3], where bounds on the

[3]Predictors in our context is often called *side information* in DSC literature.

predictors' views, $u_i(v)$ and $w_i(v)$, are:

$$
\begin{aligned}
w_i(v) &= \min(M, v + m_i) \\
u_i(v) &= \max(1, v - m_i)
\end{aligned}
\tag{1}
$$

In other words, predictor frames can be m_i views away from view v of DSC frame $F_{it+1,v}$, provided they fall within available M captured views. We assume $m_i \geq m_{i+1}$; i.e., DSC frames early in GOP may have more predictors than DSC frames later in the GOP. To be shown in Section IV, this means coding units early in GOP can enjoy strong loss protection, if deemed appropriate by an optimization.

Fig. 2 shows an example when $m_1 = 1$ and $M = 3$. By DSC's construction, as long as *one* of the predictor frames $F_{it,u_i(v)}, \ldots, F_{it,w_i(v)}$ is correctly decoded at the client's buffer, DSC frame $F_{it+1,v}$ can be correctly decoded. As shown in Fig. 2, resulting dependency among frames creates a *directed acyclic graph* (DAG). Size of a DSC frame increases with size $2m_i + 1$ of the predictor set [10]; typically, size of a DSC frame falls between a P-frame and an I-frame.

Clearly, using our proposed coding structure with DSC frames ($m_i \geq 1$) inserted, a client can switch from view v to a neighboring view $v \pm 1$ at the DSC frame boundary (by subscribing to a different WWAN multicast channel), resulting in view-switching delay of $2t/FPS$ seconds. An alternative coding structure that uses I-frames instead can also facilitate view-switching with the same delay, but requires significantly more transmission bandwidth due to the large size of I-frames.

C. WWAN Assumptions

To model WWAN packet losses, we use the Gilbert-Elliot (GE) model (a popular and commonly used model [13] for wireless losses) with independent identically distributed (iid) packet loss probabilities g and b for each of good and bad state, and state transition probabilities p and q to move between states. In other words, when a packet arrives, a weighted coin (with weight p or q depending on current state) is first tossed to determined whether it stays in the current state or transition to the other state. Then a second weighted coin (with weight g or b depending on current state) is tossed to determine if the packet is lost or not. See Fig. 3 for an illustration.
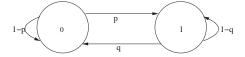

Fig. 3. Gilbert-Elliot loss model. g and b are the packet loss probabilities in 'good' and 'bad' states, respectively, and p and q are transition probabilities between states. 1 (0) indicates a bad (good) state.

Given the proposed coding structure and WWAN assumptions, we can calculate the probability α_i that a coding unit i of any view, $U_{i,\cdot}$, is *correctly received* via WWAN. We make the simplifying assumption that no more than one state transition is considered in the GE model during transmission of one unit. (For small state transition probabilities p and q, this is a good approximation.) We write α_i as:

$$
\alpha_i \approx \alpha_i^{01} + \alpha_i^{10}
\tag{2}
$$

(2) states that the probability that $U_{i,.}$ is correctly received, is the probability α_i^{01} that the network starts in good state and transits (possibly) once to bad state, plus the probability α_i^{10} that the network starts in bad state and transits to good state.

Let $n_i = s_{it+1}^{DSC} + \sum_{j=it+2}^{(i+1)t} s_j^P$ be the number of source packets in $U_{i,.}$. Let f_i be the number of FEC packets for WWAN loss protection on source packets in $U_{i,.}$. α_i^{01} can be written as:

$$\alpha_i^{01} \approx \left(\frac{q}{p+q}\right) \sum_{k=n_i}^{n_i+f_i} (1-p)^k p^{\mathbf{1}(k<n_i+f_i)} \sum_{j=n_i}^{k} C_j^k (1-g)^j g^{k-j} \tag{3}$$

where $\mathbf{1}(c)$ evaluates to 1 if clause c is true and 0 otherwise, and C_j^k is the number of combinations of k chooses j. (3) states that given transmission start in good state with probability $q/(p+q)$, if the channel remains in good state for k packets with probability $(1-p)^k$, then at least j of k packets must be correctly received for successful transmission of $U_{i,.}$. We make the simplification assumption here that all packets are lost after switching to bad state, a good assumption when b is large.

Similarly, we can derive α_i^{10} where the tranmission starts at bad state with the probability $p/(p+q)$:

$$\alpha_i^{10} \approx \left(\frac{p}{p+q}\right) \sum_{k=0}^{f_i} (1-q)^k q \sum_{j=n_i}^{n_i+f_i-k} C_j^k (1-g)^j g^{n_i+f_i-k-j} \tag{4}$$

D. Ad hoc WLAN Assumptions

We assume that the ad hoc WLAN bandwidth available for CPR is at least as large as the WWAN transmission bandwidth—a realistic assumption given larger 802.11x bandwidth relative to 3G bandwidth in practice today. Considering also that repairs are typically needed only for a fraction of the transmitted data, we do not explicitly consider a bandwidth constraint for CPR exchanges in this paper.

We assume that within a repair epoch of t/FPS seconds, state information—what packets/frames are needed by a peer—can be exchanged quickly among the peer group, so that each peer knows what packets/frames are needed by its neighbors. For an assumed small ad hoc network with only a handful of peers (to be discussed), the network diameter is small and the assumption is valid. Given these assumptions, it means also that a peer can share all helpful packets/frames he received during WWAN multicast to his neighbors via CPR during a repair epoch.

We assume on average that a peer can expect a small number N cooperative neighbors participating in CPR in an ad hoc WLAN network, each subscribing to one of M available views at a given time. For a peer watching view v, there are $b = (N-1)/M$ neighbors on average who are watching the same view. Given these b neighbors can share received packets during the repair epoch, the probability $\hat{\alpha}_i$ that a peer can correctly receive $U_{i,.}$, including CPR help from neighbors watching the same view, is:

$$\hat{\alpha}_i = \alpha_i + (1-\alpha_i)(1-(1-\alpha_i)^b) \tag{5}$$

(5) states that if a peer cannot recover his $U_{i,.}$ due to WWAN losses, then b neighbors can help repair losses if at least one of them has correctly received his $U_{i,.}$.

IV. FORMULATION

Having discussed the proposed coding structure and network models, we now formalize an optimization problem to optimize the structure parameters: size m_i of predictor set for each DSC frame $F_{it+1,.}$ of any view, and number of FEC packets f_i for each unit $U_{i,.}$. We first discuss the WWAN transmission constraint, then derive the correctly decode probability for each $U_{i,v}$. Finally, we define the objective function.

A. WWAN Transmission Constraint

We first establish the following WWAN transmission constraint: the total packets for a GOP, including source packets and FEC packets, must be less than bandwidth of C packets that corresponds to WWAN transmission of a GOP. We write the WWAN transmission rate constraint as follows:

$$\left(s_1^I + \sum_{j=2}^{t} s_j^P + f_1\right) + \sum_{i=1}^{T/t-1}\left(s_{it+1}^{DSC} + \sum_{j=it+2}^{(i+1)t} s_j^P + f_i\right) \leq C \tag{6}$$

B. Deriving Correctly Decode Probability

Given earlier defined $\hat{\alpha}_i$, we now derive the more challenging *correctly decode* probability $\beta_{i,v}$ for unit $U_{i,v}$. Due to dependencies stemming from DSC frames as shown in the DAG in Fig. 2, correct decoding of $U_{i,v}$ means there exists a *path* of correctly received coding units from $U_{i,v}$ back to unit $U_{0,.}$ of any view. We note further that unlike the DAG dependency model in [14], where correct decoding of a unit requires correct decoding of *all* its parents (an AND operator), use of DSC means that correct decoding of a unit requires correct decoding of *at least one* of its parents (an OR operator). Derivation of correctly decode probability for a DAG using an OR operator is much more involved, and is one key contribution of this paper.

Denote by (i,v) the event that $U_{i,v}$ is correctly decoded, and $\overline{(i,v)}$ its compliment. We first write $\beta_{i,v}$ as a product of the unit's correctly receive probability $\hat{\alpha}_i$, and a sum expressing the probability that at least one of its predictors is correctly decoded:

$$\beta_{i,v} = \hat{\alpha}_i \sum_{k=u_i(v)}^{w_i(v)} \gamma_{(i-1,k)\cap\overline{(i-1,u_i(v))}\cap\ldots\cap\overline{(i-1,k-1)}} \tag{7}$$

where $\gamma_{(i-1,k)\cap\overline{(i-1,u_i(v))}\cap\ldots\cap\overline{(i-1,k-1)}}$ is the probability that $U_{i-1,k}$ is correctly decoded, *and* $U_{i-1,u_i(v)}$ to $U_{i-1,k-1}$ are not correctly decoded. $\gamma_{(i-1,k)}$ with no negative event is equivalent to $\beta_{i-1,k}$. Summation in (7) essentially adds up probabilities of a sequence of disjoint events: likelihood that $U_{i-1,u_i(v)}$ is correctly decoded, *plus* likelihood that $U_{i-1,u_i(v)+1}$ is correctly decoded *and* unit $U_{i-1,u_i(v)}$ is not correctly decoded, etc. In other words, it spells out a list of disjoint events that span the event space "at least one predictor of $U_{i,v}$ is correctly decoded".

We next separate each γ in (7) into two terms, by unraveling the left-most negative clause $\overline{(i,l)}$ into two disjoint events:

either $U_{i,l}$ is not correctly received (denoted by $\overline{(i,l)}^c$), or $U_{i,l}$ is correctly received but *all* its parents are not correctly decoded (denoted by $(i,l)^c \cap \overline{(i,l)}^p$). Mathematically we write:

$$
\begin{aligned}
\gamma_{(i,v) \cap \overline{(i,l)} \cap \ldots \cap \overline{(i,v-1)}} &= \gamma_{(i,v) \cap \overline{(i,l)}^c \cap \ldots \cap \overline{(i,v-1)}} \\
&+ \gamma_{(i,v) \cap (i,l)^c \cap \overline{(i,l)}^p \cap \ldots \cap \overline{(i,v-1)}}
\end{aligned} \quad (8)
$$

Event $\overline{(i,l)}^c$ in the first term in (8) can be considered independently using the unit's correctly receive probability $\hat{\alpha}_i$:

$$
\gamma_{(i,v) \cap \overline{(i,l)}^c \cap \ldots \cap \overline{(i,v-1)}} = (1 - \hat{\alpha}_i) \, \gamma_{(i,v) \cap \overline{(i,l+1)} \cap \ldots \cap \overline{(i,v-1)}} \quad (9)
$$

$(i,l)^c$ in the second term in (8) can also be considered independently. We also replace $\overline{(i,l)}^p$ by events of $U_{i,l}$'s parents, i.e., units $U_{i-1,u_i(l)}$ to $U_{i-1,w_i(l)}$. We can now rewrite this second term $\gamma_{(i,v) \cap (i,l)^c \cap \overline{(i,l)}^p \cap \ldots \cap \overline{(i,v-1)}}$ as:

$$
= \hat{\alpha}_i \, \gamma_{(i,v) \cap \overline{(i,l+1)} \cap \ldots \cap \overline{(i,v-1)} \cap \overline{(i-1,u_i(l))} \cap \ldots \cap \overline{(i-1,w_i(l))}} \quad (10)
$$

More generally, γ may already have negative events $\overline{(i-1,.)}$'s. Unraveling of $\overline{(i,l)}^p$ will then mean writing the range of views that span the union of all negative events.

Eventually, all negative events $\overline{(i,.)}$'s are unraveled using (8) to (10), and γ is left with event (i,v) and (possibly) negative events $\overline{(i-1,.)}$'s: $\gamma_{(i,v) \cap \overline{(i-1,k_1)} \cap \ldots \cap \overline{(i-1,k_2)}}$. The key observation is that given $m_i \geq m_{i+1}$, $u_i(v) \leq k_2$; i.e., predictor set of $U_{i,v}$ overlaps $U_{i-1,k_1}, \ldots, U_{i-1,k_2}$. This is because event $\overline{(i-1,k_2)}$, $k_2 = w_i(v')$, is created when unraveling event (i,v') of a view v', and $v - v' \leq 2m_{i+1}$ (both $U_{i,v'}$ and $U_{i,v}$ are predictors of a unit U_{i+1,v^o} of some view v^o). We can hence write $\gamma_{(i,v) \cap \overline{(i-1,k_1)} \cap \ldots \cap \overline{(i-1,k_2)}}$ as:

$$
= \begin{cases} 0 & \text{if } w_i(v) \leq k_2 \\ \hat{\alpha}_i \, \sum_{k=k_2+1}^{w_i(v)} \gamma_{(i-1,k) \cap \overline{(i-1,k_1)} \cap \ldots \cap \overline{(i-1,k-1)}} & \text{o.w.} \end{cases} \quad (11)
$$

Though derivation of $\beta_{i,v}$ using (7) to (11) is quite involved, we can nevertheless bound the complexity by counting the number of γ's required for computation of all β's. Each γ contains one positive event (i,v), at most $M-1$ negative events $\overline{(i,v')} \cap \ldots \cap \overline{(i,v-1)}$ of same index i, and at most M negative events $\overline{(i-1,k_1)} \cap \ldots \cap \overline{(i-1,k_2)}$ of previous index $i-1$. Each computed γ can be stored into entry $\Gamma[i][v][v'][k_1][k_2]$ of a dynamic programming (DP) table Γ of dimension $T/t \times M \times M \times M \times M$. From (8) to (10), we see that computation of each of these Γ entries requires constant amount of operations using other Γ entries. Computation of an Γ entry using (11) requires $O(M)$ computations, but the number of such entries in Γ is $O(T/tM^3)$. Hence we can conclude that the complexity of computing all γ's is $O(T/tM^4)$.

C. Distortion-based Objective Function

We now derive expression for expected distortion in a GOP given chosen parameters for coding structure using DSC. The benefit D_i of a single unit $U_{i,.}$ of any view is the sum of the distortion reduction d_j of each frame j in the unit:

$$
D_i = \sum_{j=it+1}^{(i+1)t} d_j \quad (12)
$$

Given the earlier derived correctly decode probability $\beta_{i,v}$ of a unit $U_{i,v}$, the expected distortion D for a GOP is:

$$
D = M\bar{D} - \sum_{v=1}^{M} \sum_{i=1}^{T/t} \beta_{i,v} D_i \quad (13)
$$

where $\bar{D}$ is the initial distortion of a GOP of all views if no frames are correctly decoded.

We can now formally define our optimization problem: find structure parameters m_i's and FEC packet numbers f_i's for units $U_{i,.}$'s in a GOP that minimizes distortion in (13), subject to WWAN transmission constraint (6). We present a simple algorithm to find the optimal parameters next.

V. CODING STRUCTURE OPTIMIZATION

We search for optimal parameters for our proposed coding structure with DSC in a GOP as follows. For simplicity, assume first that our data set only has $M = 5$ captured views. Hence m_i can be one of only three values: 0, 1 or 2. Given $m_i \geq m_{i+1}$, we only need to find two DSC boundaries θ_1 and θ_2, where units $U_{i,.}$'s with $i < \theta_1$ have $m_i = 2$, units $U_{i,.}$'s with $\theta_1 \leq i < \theta_2$ have $m_i = 1$, and units $U_{i,.}$'s with $\theta_2 \leq i$ have $m_i = 0$ (P-frames). For each boundary pair (θ_1, θ_2), we do the following to find f_i's.

We first assign $\left\lfloor \frac{C - \sum_{i=0}^{T/t} n_i}{T/t * M} \right\rfloor$ FEC packets to each unit $U_{i,.}$ of any view. This ensures the WWAN transmission constraint (6) is satisfied. We then search for the best reallocation of M FEC packets from M $U_{j,.}$'s to M $U_{i,.}$'s, $j > i$, that yields the largest reduction in expected distortion (13). We iteratively perform such reallocations until no further distortion reduction is possible. These are the optimal f_i's for given DSC boundary pair (θ_1, θ_2). The optimal structure parameters are the DSC boundary pair (θ_1, θ_2) with corresponding f_i's with the lowest expected distortion.

VI. EXPERIMENTATION

To test the performance of our optimized coding structure in typical WWAN loss environment, we set up the following experiment. For source coding, we use the DSC codec in [10]— a H.263-based codec[4] with modifications to encode bitplanes of DCT coefficients given side information using low-density parity check (LDPC) codes—to encode a 100-frame MPEG multiview video test sequence akko at 320×240 resolution. We set captured views to be $M = 5$ and DSC insertion period to be $t = 10$ frames. Video playback speed is $FPS = 30$fps, meaning view-switching delay is 0.6 seconds. We fixed the quantization parameters for I-, P- and DSC frames so that the resulting visual quality in Peak Signal-to-noise Ratio (PSNR) after compression is roughly 32.5dB. Sizes of the compressed frames of the three types, r_j^I's, r_j^P's and r_j^{DSC}'s, are inputs into our structure parameter optimization. Network transport unit (NTU) is assumed to be 1250 bytes. Typical sizes for I-, P-, and DSC frames are 5, 1 or 2, and 3 packets, respectively.

---

[4]The tradeoff between DSC and FEC for evasion of error propagation and WWAN protection would be similar if a H.264-based codec is used instead.

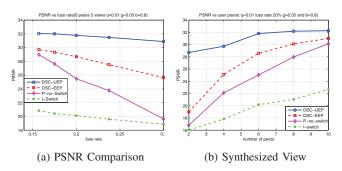(a) PSNR Comparison   (b) Synthesized View

Fig. 4. Comparison of video quality in PSNR for various coding schemes through the simulation. (a) number of cooperative peer is fixed at $N = 6$, (b) WWAN loss rate is fixed at 0.2.

For WWAN network, bandwidth for each multicast channel is assumed to be 400kbps, and packet losses were simulated according to fixed GE model parameters: $p = 0.1$, $g = 0.05$, $b = 0.8$ (q may vary to effect loss rates from 0.15 to 0.3). For ad hoc WLAN network, we set CPR bandwidth to be large enough for all peers to share all helpful packets / decoded frames.

We compare the resulting PSNR of the decoded video for four different coding structures. `DSC-UEP` is our optimized structure using DSC with UEP for a GOP. `DSC-EEP` is structure with DSC, but both m_i's and f_i's are set to constants for equal error protection for different units. `P-no-switch` is a structure with leading I- plus trailing P-frames for the entire GOP. That means view-switching every t frames is not possible. Finally, `I-switch` is a structure with I-frames inserted instead of DSC frames for view-switching. All structures exploited CPR (made possible by the available secondary network) for packet/frame recovery when possible. The question is *which* structure can best exploit bandwidth in secondary network to reap the most performance gain.

In Fig. 4(a), we see resulting PSNR of all four structures against the average loss rate of the WWAN channel (number of peers was fixed at $N = 6$); `DSC-EEP`, `P-no-switch` and `I-switch` used fixed structure for all data points, while `DSC-UEP` optimized one structure for each given network environment. We see that `DSC-UEP` is the best structure: it outperformed `I-switch` and `DSC-UEP` by 11dB and 5dB, respectively. While `I-switch` facilitated view-switching and avoided error propagation just like `DSC-UEP`, the large sizes of I-frames meant little WWAN bandwidth was left for FEC packets, f_i's, leaving the stream very vulnerable to WWAN packet losses. Notice also that `P-no-switch` outperformed `I-switch` by a wide margin when loss rate was low, but performance deteriorated quickly as the loss rate increased. This is because `P-no-switch` suffered easily from error propagation when large number of packet losses cannot be repaired via CPR.

In Fig. 4(b), we see the performance of the four structures when WWAN loss rate was fixed at 0.2 but the number of cooperative peers N varied. We see that though in general the performance improved as N increased, `DSC-UEP` outper-

formed other structures by a similarly large margin.

VII. CONCLUSION

In this paper, we address the problem of WWAN multiview video multicast, where each streaming client can interactively select one of M available views via different multicast channels as video is played back. In such scenario, packet losses are often unavoidable, and cooperation principle to exploit neighboring clients' help for packet recovery is difficult to apply. In response, we propose to optimally insert distributed source coding (DSC) frames into the encoded bitstream, so that: i) view-switching can be easily facilitated without resorting to I-frames, and ii) a neighbor watching a different view v' can nonetheless help a peer watching view v evade error propagation due to earlier packet losses, if the following DSC frame of view v is encoded using a frame in view v' as predictor. Experimental results show PSNR improvement of up to 11dB over a structure using I-frames as view-switching mechanism. It is conceivable that the same proposed DSC structure optimization can be used for WWAN multiview video unicast as well, if the unicast streams of individual views are synchronized in time.

REFERENCES

[1] J. Crowcroft and K. Paliwoda, "A multicast transport protocol," in *ACM SIGCOMM*, New York, NY, August 1988.

[2] J. Zfzal, T. Stockhammer, T. Gasiba, and W. Xu, "Video streaming over MBMS: A system design approach," in *Journal of Multimedia*, vol. 1, no.5, August 2006, pp. 25–35.

[3] X. Liu, G. Cheung, and C.-N. Chuah, "Structured network coding and cooperative wireless ad-hoc peer-to-peer repair for WWAN video broadcast," in *IEEE Transactions on Multimedia*, vol. 11, no.4, 2009.

[4] O. Alay, P. Liu, Y. Wang, E. Erkip, and S. Panwar, "Error resilient video multicast using randomized distributed space-time codes," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, March 2010.

[5] T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto, and Y. Suenaga, "Multipoint measuring system for video and sound—100 camera and microphone system," in *IEEE International Conference on Multimedia and Expo*, Toronto, Canada, July 2006.

[6] G. Cheung, A. Ortega, and N.-M. Cheung, "Generation of redundant coding structure for interactive multiview streaming," in *Seventeenth International Packet Video Workshop*, Seattle, WA, May 2009.

[7] P. Sharma, S.-J. Lee, J. Brassil, and K. Shin, "Distributed communication paradigm for wireless community networks," in *IEEE International Conference on Communications*, Seoul, Korea, May 2005.

[8] R. Bhatia, L. E. Li, H. Luo, and R. Ramjee, "ICAM: Integrated cellular and ad hoc multicast," in *IEEE Transactions on Mobile Computing*, vol. 5, no. 8, August 2006, pp. 1004–1015.

[9] G. Cheung, P. Sharma, and S. Lee, "Smart media striping over multiple burst-loss channels," in *IEEE Transactions on Selected Topics in Signal Processing*, vol. 1, no.2, August 2007, pp. 319–333.

[10] N. Cheung and A. Ortega, "Distributed source coding application to low-delay free viewpoint switching in multiview video compression," in *Proc. of Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.

[11] *Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS) user services; stage 1 (Release 6) (3GPP TS.26.246) version 6.3.0*, March 2006.

[12] *Technical Specification Group Radio Access Network; High Speed Downlink Packet Access (HSDPA); Overall description; Stage 2 (Release 10) (3GPP TS.25.308) version 10.1.0*, June 2010.

[13] I.-H. Hou and P. R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *IEEE INFOCOM*, San Diego, CA, March 2010.

[14] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," in *IEEE Transactions on Multimedia*, vol. 8, no.2, April 2006, pp. 390–404.