Modeling RPS and Evaluating Video Repair With VQM

Yubing Wang, Mark Claypool, and Robert Kinicki

Abstract—Reference Picture Selection (RPS) is a promising repair technique in lossy networks for delay-sensitive video, whereby the video encoder uses one of several previous frames as a reference frame for predictive encoding. RPS can operate in two different modes: an optimistic policy that uses negative acknowledgements (NACKs) and a more conservative policy that relies upon positive acknowledgements (ACKs). This paper compares RPS NACK and RPS ACK under various network conditions and video contents using two analytical models. The two models characterize RPS NACK and RPS ACK by incorporating the impact of reference distance on video quality, prediction dependency among video frames and Group of Pictures (GOP) length. Given packet-loss rate, round-trip time and capacity constraints, the models predict average video quality for videos with RPS ACK and RPS NACK using the Video Quality Metric (VQM). Using these two models, a series of experiments are conducted to investigate RPS performance under various conditions. The insights derived from our models can help determine appropriate choices for RPS NACK and RPS ACK under various network conditions and video contents.

Index Terms—H.264, PSNR, reference distance, repair, RPS, VQM.

I. INTRODUCTION

DESPITE many recent improvements to computer networks, streaming video quality still cannot avoid being degraded by lost data packets. A single missing video packet can propagate errors to many subsequent video frames due to inter-frame encoding dependencies. A promising repair technique for delay-sensitive video is Reference Picture Selection (RPS) [1]–[3]. In RPS, the video encoder chooses one of several previous frames as a reference frame for encoding to reduce the impact of loss. The reference frame can, by default, be the previous frame (called *RPS NACK*), or the reference frame can be several frames older if the encoder waits for receiver confirmation of successful frame reception (called RPS ACK). The performance behavior of RPS ACK and RPS NACK depends upon network latency, loss rates between the video server and client and the impact of reference distance on

the encoded video quality [4]. In particular, videos that do not significantly degrade with reference distance have the potential to benefit from RPS ACK while videos that degrade quickly with increased reference distance may prefer RPS NACK or some other repair technique.

Although numerous studies have detailed the benefits of various repair schemes to video quality [3]–[8], to the best of our knowledge, the only systematic exploration of the impact of video and network conditions on the performance of RPS ACK and RPS NACK has been done in our earlier work [9]. Our earlier work assesses RPS ACK and RPS NACK with different network losses and latencies, but relies solely on peak signal-to-noise ratio (PSNR) as the video quality metric. However this paper did not have space to present the analytic model used and provided no model validation. These shortcomings are addressed in this paper. Although widely used in the literature, PSNR has been shown to be inaccurate in predicting visual quality. This is particularly true when the video stream encounters high packet loss. However, the major focus of RPS video repair is responding to high packet loss rates.

This paper derives two analytic models to predict the quality of videos streamed with RPS NACK and RPS ACK. Rather than rely on PSNR, these models use the *Video Quality Metric* (VQM) [10], developed by the Institute for Telecommunication Sciences (ITS), to evaluate video quality. VQM provides an objective value for perceived video quality by measuring specific perceptual features of video impairment including artifacts caused by encoding methods and video quality degradation induced by network packet loss. By employing these two analytic models, this paper provides detailed analysis of RPS performance over a range of network loss and latency conditions using a variety of videos chosen to represent a diverse range in video scene complexity and motion characteristics. The basis for our video encoding model is H.264 [11], [12], an increasingly popular compression standard that supports RPS. The models incorporate a bandwidth constraint and a range of reference distances. Analysis in this paper shows that RPS performance is affected by a variety of factors including round-trip time, loss rate, video content and the Group of Pictures (GOP) length. The insights derived from the analytical models provide helpful guidelines for selecting the most effective RPS repair technique to improve video quality.

The rest of this paper is organized as follows: Section II gives background information and related work on technologies used in this paper; Section III details the analytical models for RPS; Section IV presents the experimental analysis; and Section V summarizes our conclusions and presents possible future work.
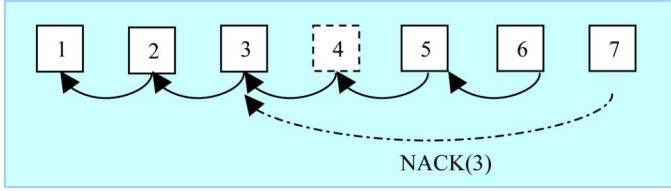
Fig. 1. Illustration of encoding using RPS with NACK Mode. Frame 4 has a transmission error.
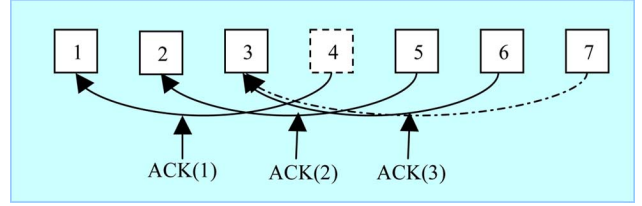

Fig. 2. Illustration of encoding using RPS with ACK Mode. Frame 4 has a transmission error.

II. BACKGROUND

A. Reference Picture Selection (RPS)

Reference Picture Selection (RPS) is a feedback-based error control technique that uses information sent by the decoder to adjust the coding parameters at the encoder to achieve better error resilience [1]–[3]. With RPS, the encoder does not always pick the previous frame, but instead selects a previously-received, correctly-decoded frame as a reference when doing predictive encoding. RPS has two modes. In RPS negative acknowledgement (NACK) mode, when there is a transmission error, the decoder sends the encoder a NACK message with the number of a previously-received, correctly-decoded frame as a reference for prediction. The encoder, upon receiving the NACK, uses the indicated correctly received frame as a reference to encode the current frame. Fig. 1 depicts RPS NACK mode. When frame 4 is determined to have a transmission error, the decoder sends a NACK to the encoder with an explicit request to use frame 3, which has been decoded correctly, for prediction. Due to network latency, the NACK arrives back at the encoder only before frame 7 is encoded. When the NACK arrives, the encoder then uses frame 3 as the reference to encode frame 7. Note, in the absence of receiving NACK messages, RPS NACK optimistically uses the most recently transmitted frame as the reference for encoding.

In RPS positive acknowledgement (ACK) mode, all correctly received frames are acknowledged and the encoder only uses acknowledged frames for references. In a scenario when there are no transmission errors, since the RPS ACK encoder uses an older reference frame than RPS NACK, the coding efficiency of RPS ACK degrades as the network round-trip time increases. However, RPS ACK mode entirely eliminates error propagation (frame 5 and 6 in Fig. 1) since the encoder only uses acknowledged frames as references. Fig. 2 depicts the use of RPS ACK mode. In this example, there are no transmission errors for the first 3 frames, allowing the encoder to receive an ACK for frame 1 while encoding frame 4. Thus, the encoder uses frame 1 as a prediction reference to encode frame 4. Similarly, the encoder uses frame 2 as a reference for frame 5, and frame 3 as a reference for frame 6. However, since no ACK is received for frame 4, frame 7 uses acknowledged frame 3, instead of frame 4, as the reference frame.

B. Impact of Reference Distance on Video Quality

By using only the successfully decoded frames as references for motion-compensated prediction, RPS ACK reduces the error propagation to subsequent frames due to data loss and achieves better error resilience. However, since the encoder may have to
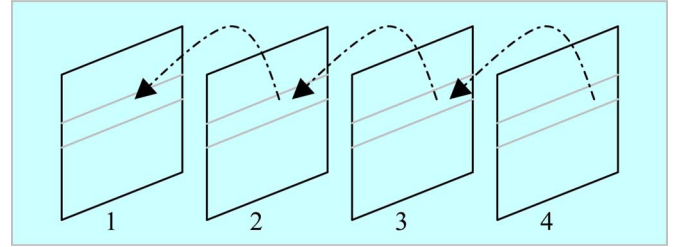

Fig. 3. Illustration of a reference chain, where each square represents a video frame, and each rectangle represents a GOB.

use an older reference frame for prediction, coding efficiency decreases as the reference distance increases since the similarity between the encoding frame and the reference frame diminishes. For a constrained network capacity, the video quality degrades as the coding efficiency drops. Our previous study quantifies video quality degradation as reference distance increases (for decoding only, without modeling repair or network loss) [4]. The degree of degradation is affected by the video content. Videos with high motion[1] tend to degrade in quality less than videos with low motion as reference distance increases. This is largely because high motion videos have a much larger number of intra-coded macro-blocks (I-blocks) and are thus less sensitive to the change of reference distance than low motion videos. Although these findings hold for both PSNR and VQM, the relationship between PSNR and reference distance can be characterized using a logarithmic function, while with VQM the same relationship can be characterized using a linear function.

III. RPS ANALYTICAL MODEL

This section derives two analytical models (i.e., one model for RPS ACK and a separate model for RPS NACK), to characterize video quality as measured by VQM. By including a variety of network characteristics that includes packet loss rate, round-trip time and bottleneck capacity constraints in the model input parameters, the models use VQM as the video quality metric to study RPS ACK and RPS NACK performance.

The models target H.264 videos since this standard incorporates both RPS ACK and NACK, but can generally represent any video encoding technique that uses RPS repair. The models assume the independent segment decoding (ISD) mode of H.264 where each Group of Blocks (GOB) is encoded as an individual sub-video independently from other GOBs in the same frame, and the reference frame is selected on a per-GOB basis, i.e., for all macro-blocks within one GOB the same reference GOB

---

[1]In this work, the amount of motion is mainly determined by the fraction of blocks that are inter-coded.

TABLE I
MODEL PARAMETERS

INPUT PARAMETERS (ENCODER)	
R_F	ENCODED FRAME RATE (IN FRAMES PER SECOND OR FPS - TYPICAL FULL-MOTION VIDEO FRAME RATES ARE 25-30 FPS)
N_G	GOP SIZE (IN FRAMES)
U_r	AVERAGE VQM FOR A GOB THAT IS ENCODED USING A REFERENCE GOB THAT IS r GOBS BACKWARD IN THE REFERENCE CHAIN
U_0	AVERAGE VQM FOR AN INTRA-CODED GOB
U'	AVERAGE VQM FOR A GOB THAT IS REPAIRED USING LOCAL CONCEALMENT
INPUT PARAMETERS (TRANSPORT)	
t_{RTT}	ROUND-TRIP TIME (IN MILLISECONDS)
P	PACKET LOSS PROBABILITY (FRACTION)
DERIVED PARAMETERS	
t_{INT}	TIME-INTERVAL BETWEEN TWO FRAMES (IN MILLISECONDS, SO 40 MSEC. FOR 25 FPS VIDEO)
q_n	PROBABILITY THAT THE N-TH GOB IN REFERENCE CHAIN IS DECODED SUCCESSFULLY
$q_{n,r}$	PROBABILITY THAT THE N-TH GOB IN THE REFERENCE CHAIN IS DECODED SUCCESSFULLY USING THE R-TH GOB AS A REFERENCE
Q_n	EXPECTED VQM VALUE FOR N-TH GOB IN THE REFERENCE CHAIN

is used. Since errors inside a GOB do not propagate to other GOBs, the video sequence can be partitioned into independent video sub-sequences. An independent video sub-sequence is referred to as a *reference chain*, illustrated in Fig. 3.

Our model assumes that each GOB is carried in a single network packet. This is a reasonable assumption since the number of GOBs per frame can be adjusted so that each GOB can be fit into a network packet. Also, reliable transmission of feedback messages are assumed since feedback is usually not part of the video syntax and is transmitted via a separate network connection where control information is exchanged [3]. The feedback connection may not suffer from congestion as does the forward link carrying the video, or may use retransmission or other methods to ensure reliable delivery. The models also assume erroneously-decoded GOBs are repaired by local concealment[2] and make no assumption on specific local concealment techniques.

A. Model Parameters

Table I provides the parameters used for our analytical models. Given the network capacity constraint and a specific video clip, the values for and U_r and U_0 are obtained from our previous work [4]. The values for U' are obtained using fixed percentages of the best value of.

B. Analytical Model for RPS ACK

RPS ACK uses acknowledged GOBs as references. Since it takes at least one round-trip time for the encoder to receive an

ACK for a GOB, the current GOB has to use a GOB which was captured at least δ [3] GOBs before it as a reference. The age of the GOB selected as a reference GOB grows linearly with the length of the round-trip time. When the encoder uses an older reference GOB, video quality is inherently lowered. As long as GOB n is successfully received, it can be decoded successfully since it can use any previously-acknowledged GOB as a reference. Therefore, the probability of GOB n being successfully decoded is

$$q_n = 1 - p. \tag{3.1}$$

Since the encoder selects the last GOB available without errors at the decoder as a reference, the reference GOB for GOB n could be chosen from GOB 1 up to GOB $(n - \delta)$. The probability of decoding GOB n correctly using GOB $(n - \delta - i)$ as a reference is:

$$(1 - p)p^i q_{n-\delta-i}, \quad 0 \le i \le n - \delta - 1, \delta = \left\lceil \frac{t_{\text{RTT}}}{t_{\text{INT}}} \right\rceil. \tag{3.2}$$

Where $q_{n-\delta-i}$ is the probability of GOB $(n-\delta-i)$ being successfully decoded, p^i is the probability of i consecutive GOBs (preceding the GOB $(n - \delta)$) having transmission errors and $(1-p)$ is the probability of GOB n being successfully received.

The use of older reference GOBs for prediction degrades the effectiveness of compression for a GOB. Thus to maintain a constant frame rate and bit rate, the encoder uses a coarser quantizer and the overall video quality may decrease. To account for video quality degradation when using an older reference GOBs for prediction, U_r denotes the average VQM for a GOB n whose reference GOB is r GOBs back in the reference chain.

The expected VQM for the n-th GOB is as follows:

$$Q_n = \begin{cases} (1-p)\sum_{i=0}^{n-\delta-1} U_{\delta+i}p^i q_{n-\delta-i} + p * U', & n > \delta \\ (1-p)U_0 + p * U', & n \le \delta \end{cases} \tag{3.3}$$

where U' denotes the average VQM for a locally concealed GOB and U_0 the average VQM for an intra-coded GOB. Note that the first δ GOBs of a GOP have to be encoded in intra mode since no ACK messages from the decoder will be received prior to encoding.

Since $q_{n-\delta-i}$ is a constant $(1-p)$, (3.3) can be further simplified as follows:

$$Q_n = \begin{cases} (1-p)^2 \sum_{i=0}^{n-\delta-1} U_{\delta+i}p^i + p * U', & n > \delta \\ (1-p)U_0 + p * U', & n \le \delta \end{cases}. \tag{3.4}$$

---

[2]Local concealment refers to repair by the decoder to hide damaged frames that cannot be repaired by other encoder-assisted means (such as RPS).

[3]$\delta = \lceil \frac{t_{\text{RTT}}}{t_{\text{INT}}} \rceil$
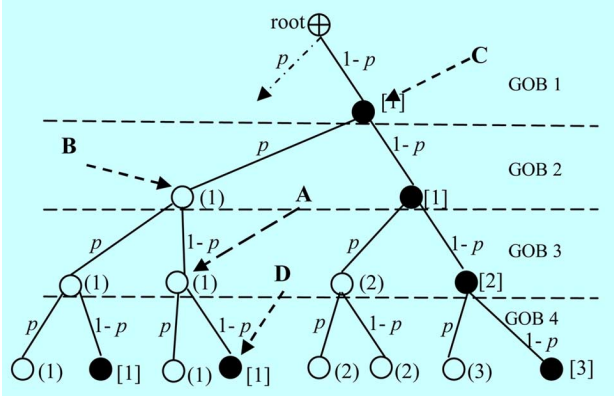
Fig. 4.   Binary Tree for the possible decoded versions of a GOB using RPS with NACK mode.

Fig. 5.   VQM versus loss with RPS NACK (video clip: Akiyo).

The average VQM over an entire GOP can be computed as follows:

$$
\begin{aligned}
Q &= \frac{1}{N_G}\left(\sum_{n=1}^{\delta}((1-p)U_0 + pU') \right. \\
&\quad \left. + \sum_{n=\delta+1}^{N_G}\left((1-p)^2\sum_{i=0}^{n-\delta-1}U_{\delta+i}p^i + pU'\right)\right) \\
&= \frac{1}{N_G}\left(((1-p)U_0 + pU')\delta \right. \\
&\quad \left. + \sum_{n=\delta+1}^{N_G}\left((1-p)^2\sum_{i=0}^{n-\delta-1}U_{\delta+i}p^i + pU'\right)\right). \quad (3.5)
\end{aligned}
$$

C. Analytical Model for RPS NACK

For RPS NACK mode, one of the GOBs in the previous frame is used as a reference GOB during the error-free transmission. After a transmission error, the decoder sends a NACK for the erroneous GOB with an explicit request to use older, intact GOBs as a reference. Therefore, the encoder may use a GOB in the previous frame or one in older frames as a reference to encode the current GOB n depending upon whether it receives a NACK from the decoder or not. If a NACK is not received from the decoder, the encoder uses a GOB in the previous frame as a reference. The probability of correctly decoding GOB n using a GOB in the previous frame as reference is denoted as $q_{n,1}$, where 1 indicates using the preceding GOB in the reference chain as a reference. If the encoder does receive a NACK, it uses the GOB requested by the decoder as a reference. As in ACK mode, the reference GOB for GOB n could be chosen from GOB 1 up to
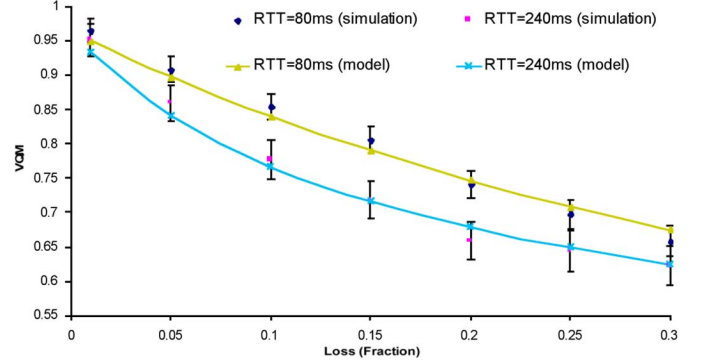
GOB $(n - \delta)$ depending upon which GOB is the last correctly decoded GOB. $q_{n,\delta+i}(0 \le i \le n-\delta-1)$ denotes the probability of decoding GOB n correctly using GOB $(n-\delta-i)$ as a reference. Since none of the first δ GOBs receives a NACK before being encoded, the successful decoding of each subsequent GOB depends upon the success of the preceding GOBs. Therefore, the probability of GOB n being successfully decoded is as follows:

$$
q_n = \begin{cases} q_{n,1} + \sum_{i=0}^{n-\delta-1} q_{n,\delta+i}, & n > \delta \\ (1-p)^n, & n \le \delta \end{cases}. \quad (3.6)
$$

The expected VQM for GOB n is shown in (3.7) at the bottom of the page. The average VQM over a GOP can be computed as shown in (3.8) at the bottom of the next page.

D. GOB Dependency Modeling

To estimate $q_{n,1}$ and $q_{n,\delta+i}(0 \le i \le n-\delta-1)$, it is essential to model the prediction dependency between GOBs in

$$
Q_n = \begin{cases} U_1 q_{n,1} + \sum_{i=0}^{n-\delta-1} U_{\delta+i} q_{n,\delta+i} + (1-q_n)U', & n > \delta \\ (1-p)^n U_1 + (1 - (1-p)^n)U', & 1 < n \le \delta \\ (1-p)*U_0 + p*)U', & n = 1 \end{cases}. \quad (3.7)
$$

the reference chain. A binary tree is used to model GOB dependency for RPS with NACK mode. Two input parameters are required to build the dependency tree: packet loss probability (p) and round-trip time (δ). Fig. 4 illustrates a binary tree for the possible decoded versions of a GOB and the corresponding reference GOB selections while using RPS with NACK mode. In the illustrated example, there are four GOBs and the round-trip time equals the length of time to capture two GOBs.

A node in the tree represents a decoded version of a GOB in a video frame. The nodes with hollow circles are those decoded erroneously while those with solid circle are those decoded correctly. Branches leaving a node represent the two cases that either a packet[4] is received erroneously with probability p or received correctly with probability $(1 - p)$. For a node whose decoded status is "Erroneous", the number (in parenthesis) besides the node represents the last GOB that has been decoded correctly at the decoder when entering this node; for a node with "Correct" decoded status, the number (in brackets) represents its reference GOB number. The GOB number between two dashed lines represents the GOB in transmission. Note that the root node (labeled with a crossed circle) represents an intra-coded GOB. Therefore, each time the encoder intra-codes a GOB, the binary tree will be refreshed.

To illustrate how the decoded status of a node is decided, consider node A (GOB 3). Since the ancestor of node A (node C) is received correctly, node A (GOB 3) did not receive a NACK message from the decoder and thus used its parent (node B, GOB 2) as a reference. Since the decoded status of node B is "Erroneous," the decoded status of node A is "Erroneous" as well. Upon entering node A, the last correctly decoded GOB at the decoder is GOB 1, therefore, the number (in parenthesis) besides node A is 1. Next, consider node D (GOB 4). Since D's ancestor 2 frames back (node B, GOB 2) is received erroneously (under the p branch), node B receives a NACK message from the decoder, which explicitly requires using GOB 1 as a reference. Thus, the reference GOB for node B is GOB 1. Furthermore, node B is under the $(1 - p)$ branch to indicate GOB 4 was correctly received. Therefore, the decoded status of node B is "Correct."

E. GOB Dependency Tree Creation

Each node in the GOB dependency tree contains the following information.
- *Decoded Status*: Correct or Erroneous.
- Probability of occurrence of this node (decoded version of a GOB).
- *Latest GOB that has been decoded correctly at the decoder when entering this node* (LDC[5] for short) - this informa-

[4]As stated earlier, our model assumes each packet contains one GOB.

[5]LDC stands for Latest Decoded Correctly

tion is recorded only for a node with "Erroneous" decoded status.
- *Reference GOB* - this information is recorded only for a node with "Correct" decoded status.
- GOB number in the reference chain.

The creation of the GOB dependency tree begins with a successfully decoded intra GOB. Since no correctly decoded GOBs are under the p branch of the root node, that part of the tree can be ignored. For each node, the four parameters described above are determined using the following algorithm.

1. If this node is under the p branch:
   - Its decoded status is set to "Erroneous";
   - Its probability is set to: $p * \text{parent} \rightarrow \text{prob}$, where $\text{parent} \rightarrow \text{prob}$ is its parent's probability;
   - Its LDC is based upon its parent's decoded status. If its parent is decoded correctly, its LDC is set to its parent's GOB number; otherwise the LDC is set to its parent's LDC;
   - Its GOB number is set to its parent's GOB $\text{number} + 1$.

2. If this node is under the $(1 - p)$ branch:
   - Its decoded status is based upon whether its ancestor δ frames back is received correctly and the decoded status of its parent. If its ancestor δ frames back was received correctly (no NACK), its parent's decoded status is checked. If its parent is decoded correctly, its decoded status is set to "Correct", otherwise it is set to "Erroneous." If its ancestor δ frames back was *not* received correctly, its decoded status is set to "Correct" since it received a NACK message from the decoder and used an older, correctly decoded GOB as a reference.
   - Its probability is set to: $(1 - p) * \text{parent} \rightarrow \text{prob}$, where $\text{parent} \rightarrow \text{prob}$ is its parent's probability;
   - If its decoded status is "Correct", its reference GOB is determined based upon the decoded status of its ancestor δ frames back. If its ancestor δ frames back was received correctly (no NACK), its reference GOB is set to its parent's GOB number. If its ancestor δ frames back was NOT received correctly, its reference GOB is set to its ancestor's LDC.
   - If its decoded status is "Erroneous", its LDC is determined based upon the decoded status of its parent. If its parent was decoded correctly, its LDC is set to its parent's GOB number; otherwise it is set to its parent's LDC.
   - Its GOB number is set to its parent's GOB number $+1$.

F. Estimate $q_{n,r}$ With the GOB Dependency Tree

After building the GOB dependency tree, $q_{n,r}$ is estimated in two steps. First, the GOB dependency tree is traversed to find all "Correct" nodes with GOB number equal n and reference GOB number $(n - r)$. Then the probabilities of each node from step 1 are added together to produce an estimate for $q_{n,r}$.

$$Q = \frac{1}{N_G} \left(\begin{array}{c} \sum_{n=1}^{\delta} ((1-p)^n U_1 + (1 - (1-p)^n)U') + \\ \sum_{n=\delta+1}^{N_G} (U_1 q_{n,1} + \sum_{i=0}^{n-\delta-1} U_{\delta+i} q_{n,\delta+i} + (1 - q_n)U') \end{array} \right). \tag{3.8}$$

Fig. 6.   VQM versus RTT with RPS NACK under different loss rates.

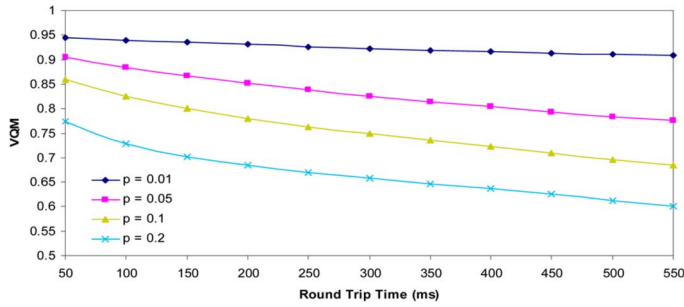Fig. 8.   VQM versus Loss with RPS NACK under different loss rates.

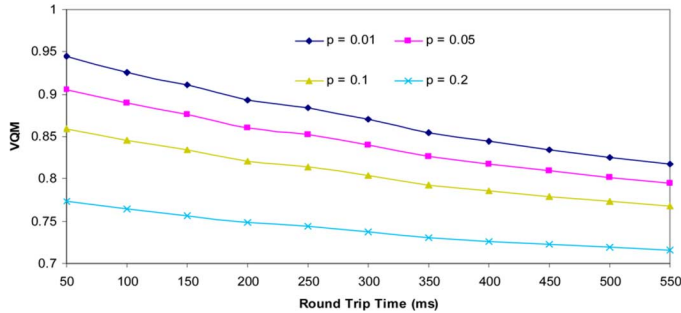Fig. 7.   VQM versus RTT with RPS ACK under different loss rates.

Fig. 9.   VQM versus loss with RPS ACK under different loss rates.

IV. MODEL VALIDATION

Simulation experiments are designed to verify the accuracy of using these analytical models in predicting video quality under various network conditions. Comparing performance predicted by the analytical models against simulated performance can provide an indication of the accuracy of our analytical models. To conserve space, only the simulations for RPS NACK are presented but the RPS ACK results are similar.

The simulations modify the input video sequences based on the given loss probability and round-trip delay to mimic the effect of packet loss as well as the change of reference distance imposed by RPS NACK on the video quality. The modified input sequences are encoded using H.264 and the average video quality in terms of VQM is measured and compared against that predicted by our analytical models.

In order to simulate the action of the H.264 encoder in response to frame loss, the encoder needs to be able to alter the reference distance in a controllable manner while encoding a video sequence. One way to achieve this is to modify the H.264 encoder to select the reference pictures specified by users. However, this approach is complicated as it involves modifying the encoder and may result in inaccurate measurements if done incorrectly. We take an alternative approach by changing the input video sequences as described later in this section, resulting in a decreased chance of human-induced error with comparable fidelity.

The simulations vary round-trip time, loss probability as well as video contents. For a specific video clip, four input sequences with different loss patterns are generated based on the given loss probability and round-trip delay. Each modified input sequence is encoded using H.264 and the VQM for each encoded video sequence is measured and compared against the VQM value predicted by the analytical model. VQM is not built into the
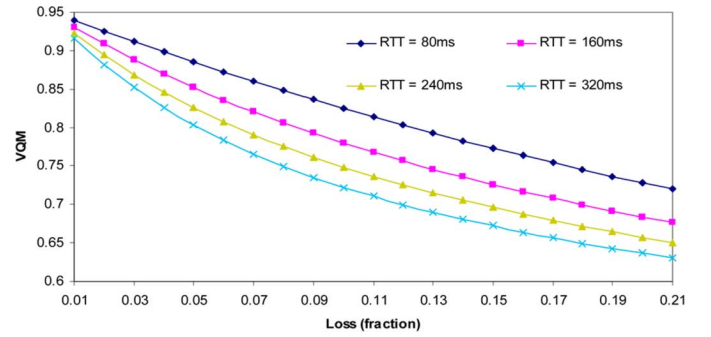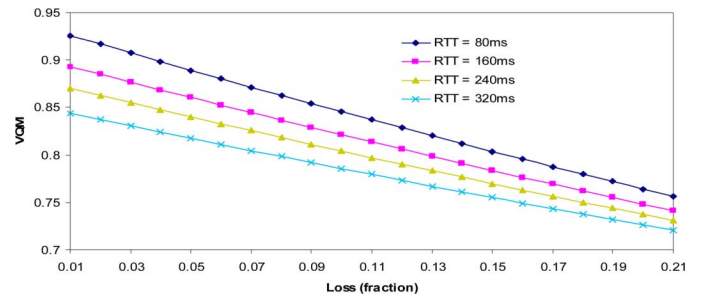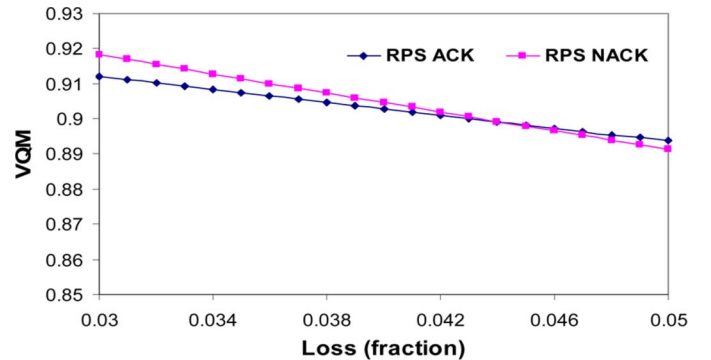

Fig. 10.   RPS NACK versus ACK with RTT 80 ms.

H.264 encoder, so a VQM measurement tool named *VQM-PC* is used[6]. This VQM tool takes the original and the processed video clips as input and measures the video quality of the processed video clips relative to the original video clips. The resulting VQM score is in the range of (0, 1), where 0 represents no impairment and 1 represents the maximum impairment. For better comparisons with PSNR (i.e., higher values are better), (1-VQM) is used as the quality metric so that a 1 represents the best quality and 0 represents the worst.

Fig. 5 shows the average VQM values for the simulations along with the VQM values predicted by the analytical model. The video clip for this experiment is *Akiyo* which is a low motion video clip. The loss probability is varied from one percent up to thirty percent and the round-trip time is varied from 80 ms to 240 ms. For the given loss probability and round-trip delay, the average VQM over four simulation experiments is presented along with its 95 percent confidence interval shown with an error
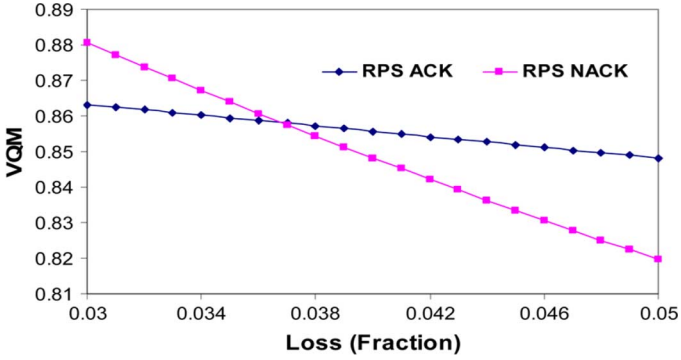
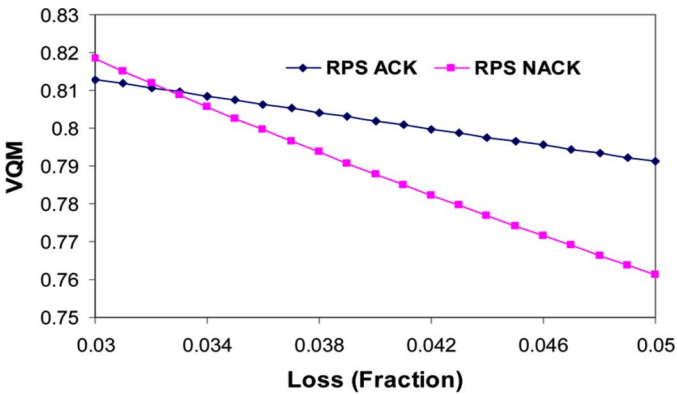Fig. 11. RPS NACK versus ACK with RTT 160 ms.

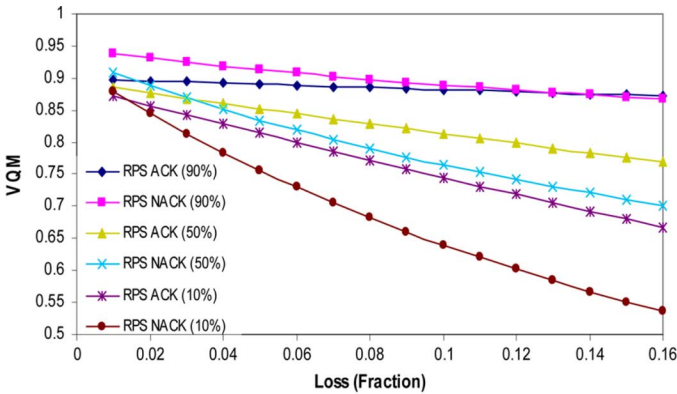Fig. 12. RPS NACK versus ACK with RTT 400 ms.

Fig. 13. RPS ACK versus RPS NACK by varying quality for locally concealed GOBs.

bar. The two curves illustrate the VQM values predicted by the analytical model under varying loss probability and round-trip delay. Both curves are within the 95% confidence intervals of all simulation samples, indicating the VQM values predicted by the analytical model are consistent with the simulations results. As the loss probability increases, the variance is also increased from 0.018 to 0.0.22 for 80 ms round trip delay and 0.024 to 0.028 for 240 ms round trip delay. Raising the round-trip delay increases the variance. This is expected since increased round-trip time produces longer sequences of error propagation.

V. ANALYSIS

This section uses the analytic models presented in the previous section to analyze RPS performance over a range of network conditions. The analytic experiments select a set of video clips with a variety of motion content. Each video sequence contains 300 video frames captured at a rate of 25 frames/s (fps). The content of these video clips can be categorized into one of three approximate groups based on motion/scene complexity: high, medium and low. Table II provides the approximate content classification for each video clip, with an identifying name and a brief description of the video content.

The impact of round-trip time on RPS video quality is examined first. Fig. 6 depicts VQM versus round-trip time for one video (*News*) encoded with RPS NACK and GOP size 22, under four loss rates ranging from 1% to 20%. The quality for a locally concealed GOB is 50% of the best quality of a GOB. As round-trip time increases, in Fig. 6, the video quality (VQM) degrades for all loss rates. However, the amount of quality degradation is not uniform. With RPS NACK, video quality under the higher packet loss rates degrades faster with an increase in round-trip time than under lower packet loss rates. For RPS NACK, each transmission error propagates to subsequent frames for a period of one round-trip time. Thus, higher packet loss rates induce more frequent GOB error propagation and video quality degrades more quickly with higher round-trip times. Fig. 7 depicts VQM versus round-trip time for a News video encoded with RPS ACK under the same four loss rates. As the round-trip time increases the average VQM for videos with RPS ACK degrades for all loss rates, similar to RPS NACK. However, unlike with RPS NACK, RPS ACK video quality degrades slower with an increase in round-trip time and higher packet loss rates than under lower packet loss rates. When the packet loss rate is low, the major cause of video quality degradation for RPS ACK is the increased reference distance caused by the round-trip time; whereas under higher packet loss rates, the video quality degradation for RPS ACK is attributed more to packet loss than to round-trip time.

The impact of loss rate on RPS video quality is investigated next. Fig. 8 provides VQM versus loss rate curves for the *News* video encoded with RPS NACK for four round-trip times ranging from 80 ms to 320 ms. As the loss rate increases, the video quality (VQM) using RPS NACK degrades for all round-trip times. However, with RPS NACK, video quality under higher round-trip times degrades faster than does the video quality under lower round-trip times. With RPS NACK, larger round-trip time implies longer error propagation periods that cause video quality to degrade more rapidly. Fig. 9 graphs VQM versus loss probability curves for a video encoded with RPS ACK for the same four round-trip times. As loss rate increases, like RPS NACK, for RPS ACK the video quality degrades for all round-trip times. However, unlike RPS NACK, for RPS ACK, the video quality under higher round-trip times degrades slower than those under lower round-trip times. Under higher round-trip times, video quality degradation for RPS ACK is attributed more to the round-trip time than to the packet loss, whereas under lower round-trip times, packet loss is the dominant cause of video quality degradation.

Similar trends were observed for other tested video clips representing a variety of motion characteristics, but these results are omitted to avoid redundancy.

Analysis thus far demonstrates that video quality for RPS ACK and RPS NACK is affected by round-trip time and packet loss rate. To make an informed choice about RPS, it is useful to know the range of packet loss within which RPS NACK performs better than RPS ACK and vise versa, and how this relationship changes with round-trip time, local concealment and video content. Figs. 10–12 compare RPS ACK and RPS NACK by graphing VQM versus packet loss with each figure having a different round-trip time. All three experiments again use the *News* video clip (the effects of video content is analyzed later). As shown in Fig. 10 with an 80 ms round-trip time, when the loss rate is less than 0.044, RPS NACK outperforms RPS ACK and when the loss rate is larger than 0.044, RPS ACK performs better than RPS NACK. When round-trip time is increased from 80 ms to 160 ms in Fig. 11, the same crossover point is reduced from 0.044 to 0.037. In Fig. 12 with a 400 ms round-trip time, the crossover point is further reduced to 0.032. This confirms that as round-trip time increases, the video quality with RPS NACK degrades faster than RPS ACK. For RPS NACK, increased round-trip time produces longer GOB error propagation; whereas for RPS ACK, increased round-trip time yields higher GOB reference distances. Increasing error propagation does more harm to video quality than does increasing reference distance.

How local concealment affects the crossover point is investigated further. Fig. 13 shows the crossover points when the quality for a locally concealed GOB is 90%, 50% and 10% of the best quality of a GOB respectively with a round-trip time of 160 ms. When the quality for a locally concealed GOB is 90% of the best quality of a GOB, the packet loss crossover point is 0.13. When the locally concealed quality is reduced to 50% of the best quality of a GOB, the crossover point is reduced to 0.037 and further reduced to 0.01 when the locally concealed quality is reduced to 10% of the best quality of a GOB. This suggests that RPS NACK outperforms RPS ACK over a wider range when there is better local concealment.

The relationship between crossover point and round-trip time for different video content is investigated next. Fig. 14 shows the quality crossover point versus round-trip time for the six videos in Table II. For loss rates above the trend-lines, RPS ACK performs better than RPS NACK while for loss rates below the trend-lines, RPS NACK performs better than RPS ACK. As round-trip time is increased, the crossover points are lowered for all videos. This suggests that regardless of the video content, increasing the error propagation is more harmful to video quality than increasing reference distance. For a fixed round-trip time, the crossover points for low-motion videos are higher than the crossover points for high-motion videos. This implies that RPS ACK outperform RPS NACK over a wider range of packet loss rates for high-motion videos than for low-motion videos. High-motion videos are less sensitive to changes in reference distance and thus can achieve better video quality with RPS ACK than can low-motion videos.

Similar trends were observed for videos using PSNR as the quality metric [9]. Fig. 15 shows crossover point versus round-
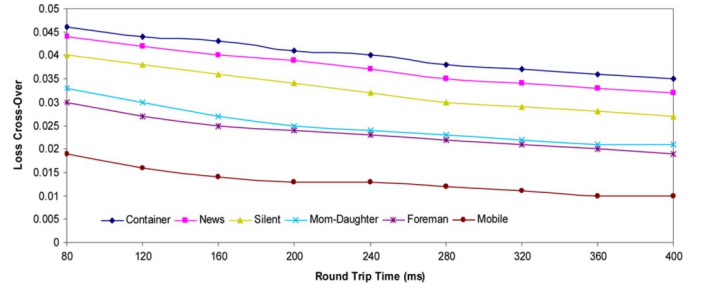

Fig. 14.   Loss cross-over point for loss versus round-trip time for six video clips using VQM.
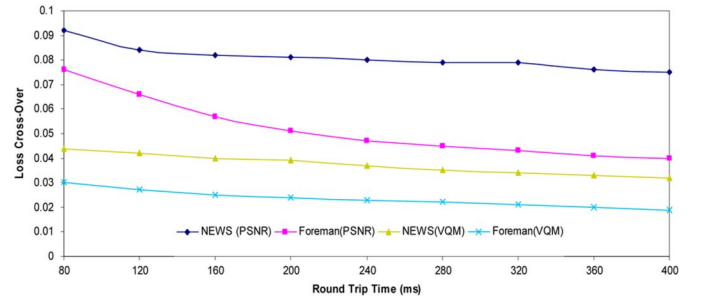

Fig. 15.   Loss cross-over point for loss versus round-trip time for two videos using both PSNR and VQM.

trip time for two videos evaluated with both PSNR and VQM for quality metrics. Fig. 15 clearly shows that for both videos the crossover points when using PSNR to measure quality are higher than when using VQM to measure quality. For instance, for the *News* video, when the round-trip time is 200 ms, the cross-point for PSNR is 0.085, whereas for VQM, the cross-point is 0.04. Hence, the range of loss rates where RPS NACK outperforms RPS ACK is smaller when using VQM to predict quality than when using PSNR to predict quality. This implies that VQM, a metric designed to incorporate both temporal and spatial aspects of video, is more sensitive to loss than PSNR, a metric that captures only spatial degradations in video.

Finally, the RPS models are used to investigate the impact of the GOP length on video quality. Fig. 16 depicts VQM versus GOP length for a video encoded with RPS NACK for round-trip times ranging from 80 ms to 320 ms, as well as a video with no repair. The loss rate for this experiment is 0.05. Below a GOP length of 5, video quality increases in all cases. After the GOP length reaches 5, quality for the video without RPS degrades due to error propagation. With RPS NACK, when the round-trip time is 80 ms, quality increases and becomes asymptotically steady. When round-trip times are 160 ms, 240 ms and 320 ms, quality slightly decreases and becomes asymptotically steady. For all GOP lengths, videos with RPS NACK perform no worse than videos without RPS, and RPS NACK performs better for lower round-trip times than for higher round-trip times since higher round-trip times introduce longer periods of error propagation. The results shown in Fig. 16 can be compared with those in [13], which studied the impact of the choice of GOP on video quality for MPEG video and Forward Error Correction (FEC) repair. The results in Fig. 16, that GOP lengths larger than 5 have diminishing returns, are similar to the result in [13]
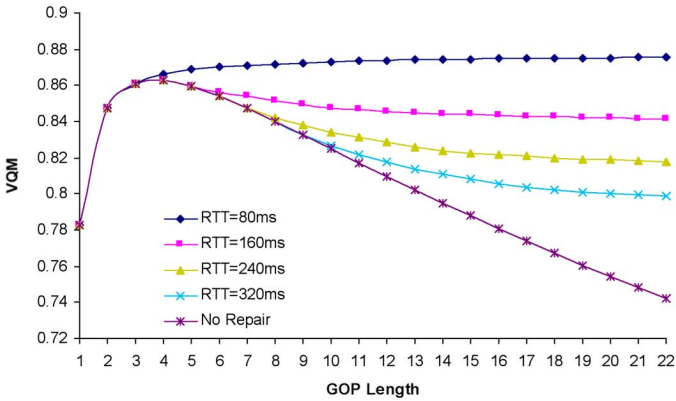
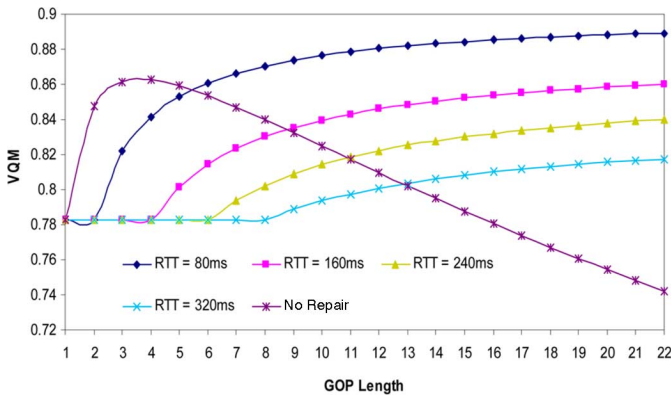Fig. 16. VQM versus GOP length with RPS NACK.

Fig. 17. VQM versus GOP length with RPS ACK.

despite different video encoding (H.264 versus MPEG) and different repair methods (RPS versus FEC).

Fig. 17 depicts VQM versus GOP length for a video encoded with RPS ACK for the same four round-trip times. As GOP length increases, quality increases for videos with RPS ACK for all round-trip times shown. Since RPS ACK uses intra-coding before any frames are acknowledged, the quality for the first part of the GOP remains constant and increases only after ACKs are received by the encoder. For all GOP lengths, RPS ACK performs better under lower round-trip times than under higher round-trip times since RPS ACK under higher round-trip times has to use older frames as references for prediction. It is worth noting that below a certain GOP size, videos without repair perform better than videos with RPS ACK for repair. For instance, when the round-trip time is 80 ms and GOP length is below 5, videos without RPS have better quality than videos with RPS ACK. Videos without RPS always use the previous GOB as reference and rely on intra coding to stop error propagation. When the GOP length is small, error propagation can be quickly stopped, whereas RPS ACK always uses older frames as references. Therefore, when the loss rate is low and the GOP length is small, videos without RPS outperform videos with RPS ACK.

VI. CONCLUSION AND FUTURE WORK

This paper compares RPS NACK and RPS ACK under various network conditions and video content using two analytical

models. Our models characterize RPS NACK and RPS ACK, incorporating the impact of reference distance on video quality, prediction dependency among GOBs in the reference chain and Group of Picture (GOP) length. Given a variety of network characteristics including packet-loss rate, round-trip time, capacity constraints, and measured video quality derived from earlier empirical studies, the models predict average video quality for videos with RPS ACK and RPS NACK. Validation through simulation suggests the models accurately predict video quality.

Analytic experiments over a range of loss rates, round-trip times and video content using the models show the quality of video with RPS ACK is more sensitive to round-trip times, whereas the quality of video with RPS NACK is more sensitive to packet loss. For a given round-trip time, the loss rate where RPS ACK performs better than RPS NACK is higher for low motion videos than it is for high motion videos. Videos with RPS NACK perform no worse than videos without RPS for all GOP lengths. However, below certain GOP lengths, videos without RPS perform better than videos with RPS ACK. For some small GOP lengths and high round-trip times, videos with RPS ACK actually perform worse than videos with no repair. In general, better methods of local concealment increase the range where RPS NACK outperforms RPS ACK.
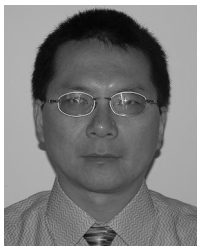
Rather than rely upon PSNR, a metric that captures only spatial degradations in video, our models predict video quality using VQM, a metric designed to incorporate temporal as well as spatial aspects of video. Overall, the experimental results using VQM yield trends similar to results when PSNR is used, except the experiments presented indicate that VQM is more sensitive to network loss than PSNR.

There are two other popular feedback based error control techniques: selective retransmission [14] and intra update [3]. Future work could develop analytical models for these two error control techniques and compare them with RPS under various network conditions and video content. Future work could also do further evaluation comparing RPS to alternative repair techniques under additional bandwidth and other network constraints, as well as investigate scalable coding in conjunction with RPS. Other future work could build a videoconference system that automatically adapts to the best RPS mode (ACK or NACK) depending upon the network conditions and video content.

REFERENCES

[1] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K., Nov. 1996, vol. 3, pp. 1503–1508.

[2] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified interframe coding system for limited reference picture memories," in *Proc. Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp. 743–748.

[3] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE, Special Issue on Video for Mobile Multi-media*, vol. 97, no. 10, pp. 1707–1723, Oct. 1999.

[4] Y. Wang, M. Claypool, and R. Kinicki, "Impact of reference distance for motion compensation prediction on video quality," in *Proc. ACM/SPIE Multimedia Computing and Networking (MMCN) Conf.*, San Jose, CA, Jan. 2007.

[5] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE, Special Issue on Video for Mobile Multi-media*, vol. 86, no. 5, May 1998.

[6] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, pp. 61–82, July 2000.

[7] K. Mayer-Patel, L. Le, and G. Carle, "An MPEG performance model and its application to adaptive forward error correction," in *Proc. ACM Multimedia*, Juan-les-Pins, France, Dec. 2002.

[8] H. Wu, M. Claypool, and R. Kinicki, "A model for MPEG with forward error correction and TCP-friendly bandwidth," in *Proc. Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Monterey, CA, Jun. 2003.

[9] Y. Wang, M. Claypool, and R. Kinicki, "An analytic comparison of RPS video repair," in *Proc. ACM/SPIE Multimedia Computing and Networking (MMCN) Conference*, San Jose, CA, Jan. 2008.

[10] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 312–322, Sept. 2004.

[11] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*.   New York: Wiley, 2004, ISBN 0-470-84837-5.

[12] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC) May 2003, Joint Video Team of ITU-T and ISO/IEC JTC 1, Doc. JVT-G050r1.

[13] H. Wu, M. Claypool, and R. Kinicki, "Guidelines for selecting practical MPEG group of pictures," in *Proc. IASTED Int. Conf. Internet and Multimedia Systems and Applications (EuroIMSA)*, Innsbruck, Austria, Feb. 2006.

[14] I. Rhee, "Error control techniques for interactive low-bit rate video transmission over the internet," in *Proc. ACM SIGCOM*, Vancouver, BC, Canada, Sep. 1998.

Mark Claypool received the M.S. and Ph.D. degrees from the University of Minnesota, Minneapolis, in 1993 and 1997, respectively.

He joined the Computer Science department at the Worcester Polytechnic Institute (WPI), Worcester, MA, in 1997, receiving tenure and promotion to Associate Professor in 2004. He is also the Director of the Interactive Media and Game Development major at WPI, a four-year degree in the principles of interactive applications and computer-based game development. His primary research interests include multimedia networking, congestion control, and network games.

Robert Kinicki graduated from Duke University, Durham, NC, in 1978.

He has been a faculty member at Worcester Polytechnic Institute (WPI), Worcester, MA, since 1978. His research interests include performance of wireless sensor networks and wireless networks, multimedia applications and congestion control.

Yubing Wang received the M.S. and Ph.D. degrees in computer science from Worcester Polytechnic Institute (WPI), Worcester, MA, in 2002 and 2008.

He has been with the EMC Corporation, Southborough, MA, for the past eight years, participating in the developments of several EMC NAS products. His primary research interests include multimedia networking, error control for video transfer, and real-time operating systems.