SCALABLE CROSS-LAYER RATE ALLOCATION FOR IMAGE TRANSMISSION OVER HETEROGENEOUS WIRELESS NETWORKS

Wei Yu, K. J. Ray Liu
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD, 20742

Zoltan Safar
Department of Innovation
IT University of Copenhagen
Copenhagen, Denmark

ABSTRACT

In this paper, we present an efficient real-time JPEG 2000 image transmission system over heterogeneous wireless networks. By exploiting the advantages of UDP-lite protocol and cross-layer optimization, the system can utilize the limited bandwidth more efficiently. Moreover, a scalable cross-layer rate allocation (SCRA) algorithm is proposed to perform the cross-layer optimization. The algorithm can manage a large number of adjustable parameters with low complexity. The simulation results show that by using the proposed method, 1.5dB gain in PSNR can be obtained for a fixed transmission rate of 0.5bpp, or 30% transmission time can be saved under an end-to-end distortion constraint of 30dB in PSNR.

1. INTRODUCTION

With the explosive growth of the Internet and dramatic increase in wireless access, there is a tremendous demand for multimedia delivery over wireless Internet. However, due to severe wireless channel conditions, multimedia delivery over wireless Internet is more difficult than over wired Internet. In case of real-time image transmission over heterogeneous wireless packet networks, the data is transmitted packet by packet, and the transmission path includes both wired and wireless links. The designers of efficient image transmission systems over such heterogeneous wireless networks must face many challenges.

One major challenge is the high bit error rate (BER) and the inefficiency of the traditional packet dropping scheme for multimedia delivery. In a traditional packet-based network, besides network congestion, any residual errors will cause a packet to be dropped. Although this packet dropping scheme works well for reliable data communications, it is not efficient for real-time wireless multimedia transmission. First, multimedia data is usually *error tolerant*, so even if some errors are introduced, the original information may still be reconstructed with tolerable distortion. Second, due to severe channel conditions, wireless channels usually exhibit high BER. Even with some channel coding techniques, the packet dropping rate (PDR) is still very high. Third, the real-time constraint excludes packet retransmission in case of packet drops. To overcome these problems, in addition to applying channel coding to reduce the residual BER at the data link layer and to combat packet erasure at the IP or upper layers, more efficient dropping strategy should be applied for multimedia delivery, such as UDP-lite. This protocol only performs error check on certain part of the received packet, and leaves error handling of the remaining data to the upper layers [1]. Moreover, a wireless link may utilize multiple transmit and receive antennas to reduce the BER and PDR, and OFDM modulation to combat inter-symbol interference.

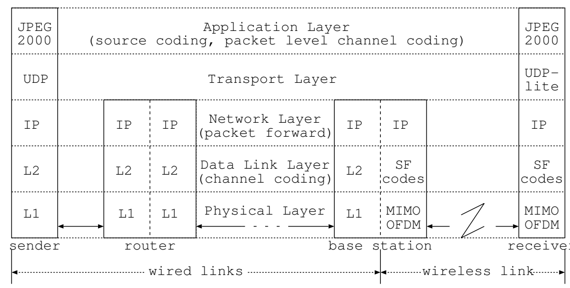

Fig. 1. Network model for the C-JSCC system

Another major challenge comes from the low-bandwidth and time-varying characteristic of wireless channels. Due to the terminal's mobility, interference, and fading, the wireless channel usually exhibits low bandwidth, and channel conditions may change dramatically from time to time. To alleviate these effects, the image coding and transmission strategies should be adaptively adjusted to match current channel conditions, and joint source channel coding (JSCC) should also be applied to exploit the special characteristics of multimedia data, such as *unequal importance*.

In this paper, we propose a real-time image transmission system for heterogeneous wireless packet networks. The objective is to adjust the image coding and transmission strategies adaptively to achieve the best resource utilization. Space-frequency (SF) coded MIMO-OFDM system is used in the physical layer of the wireless link [2], and the source coding method is JPEG 2000 [3]. Unlike traditional systems which only perform JSCC at application layer, cross-layer JSCC is performed to improve the performance of the proposed system. By adopting UDP-lite and performing cross-layer optimization, the limited resources can be distributed across network layers to achieve better efficiency. In the proposed system, cross-layer optimization is performed through the *Scalable Cross-layer Rate Allocation* (SCRA) algorithm, which can manage the large number of adjustable parameters with low complexity and can still approximate the optimal solution well.

2. SYSTEM DESCRIPTION

2.1. System Description

The network model and protocol stacks employed in the proposed cross-layer JSCC (C-JSCC) system is illustrated in Fig. 1. In this paper, we focus on point-to-point transmission, and assume only the first or last link is wireless, which is a common situation in heterogeneous wireless networks. Without loss of generality, we assume that the receiver is a wireless mobile, and call the node directly connected to the mobile as wireless base station. In the C-JSCC system, UDP-lite is used only at the wireless end of the link, based on the reasonable assumption that packet dropping in the wired network is only caused by network congestion.

Real-time delivery of images in the C-JSCC system is carried out with the cooperation of several network layers. Source coding and packet-level channel coding is performed at the application layer (APL), bit level channel coding at the data link layer (DLL) and physical layer (PHY). If UDP-lite is used, interleaving and bit-level channel coding can also be applied at APL layer to deal with the residual errors in the received packet, which can reduce the complexity of the DLL layer implementation and improve the performance of APL layer channel coding.

JPEG 2000 is adopted as the source coding standard in the C-JSCC system, since it has the following desirable properties: high compression efficiency, quality scalable compression, and strong error resilience, which are of importance for image transmission over error-prone channels. The JPEG 2000 encoder can work in two modes: compression and transcoding. In the compression mode, it generates a quality scalable bitstream with multiple quality segments in the rate-distortion (RD) optimal sense. In the transcoding mode, the goal is to make the bitstream more robust by applying various forms of error resilient coding schemes. As a result, an image can be compressed one time and transcoded many times for different channel conditions.

To combat inter-symbol interference and improve the wireless link quality by exploiting spatial and frequency diversity, MIMO-OFDM system is adopted at the PHY layer and variable rate SF are codes used [2] at the DLL layer. The advantage of the variable rate SF codes lies in its low complexity, since redundancy is introduced through repetition. Two types of codes are employed at APL layer: packet erasure (PE) codes [4] to combat packet drops, and Reed Solomon (RS) codes [5] plus interleaving to combat residual errors in the received packets.

In the C-JSCC system, channel conditions are first estimated before each image is transmitted. Given the estimated channel conditions and performance requirement (maximum transmission time or distortion), a quality scalable bitstream with multiple segments is first generated using JPEG 2000 in the compression mode. Based on the channel conditions and characteristics of image content, the SCRA algorithm is applied to find the best parameters across network layers such that the end-to-end distortion is minimized under transmission time constraint, or the transmission time is minimized under distortion constraint. The adjustable parameters include the types of error resilient source coding schemes, the number of quality segments to be transmitted, and the channel coding rates at different network layers.

2.2. Problem Formulation

In the C-JSCC system, each segment is transmitted using several IP packets, and PE codes are applied inside quality segments. At the receiver, after all the channel decoding, two types of errors may still remain, segment erasure and random bit errors inside segments. In JPEG 2000, if one quality segment is erased, then all the following quality segments will not be used to reconstruct the image, while if some bit errors happen inside a quality segment, the segment can still be used by the source decoder, and the error effects are determined by the error propagation range and the weighting factors of the affected bits, where the weighting factor of one bit denotes the distortion introduced if this bit gets corrupted.

We assume the channel conditions do not change during the transmission of each image. For each quality segment l of the compressed bitstream, let S_l^s, S_l^f, S_l^r, and S_l^p denote the set of source error resilient coding schemes, SF code rates, RS code rates, and PE code rates, respectively, and let $s_l = \langle s_l^s, s_l^f, s_l^r, s_l^p \rangle$ denote the coding and transmission strategy for this segment, where $s_l^s \in S_l^s$, $s_l^f \in S_l^f$, $s_l^r \in S_l^r$ and $s_l^p \in S_l^p$. Let s_l^0 denote that the segment l is not transmitted, and $\bar{s} = \langle s_1, s_2, \ldots, s_L \rangle$ the coding and transmission strategy for a whole image.

Let D_{total} denote the overall distortion if no segments are received, and $G_l(s_l)$ the expected gain (distortion reduction) when strategy s_l is applied on segment l and no segment with index less than l has been dropped, $P_{drop}(l, \bar{s})$ denote the probability that at least one segment with index less than l has been erased by the channel, which depends only on strategies from s_1 to s_{l-1}. Then the overall expected gain $G(\bar{s})$ and the overall expected end-to-end distortion $D(\bar{s})$ can be approximated as

$$G(\bar{s}) = \sum_{l=1}^{L} (1 - P_{drop}(l, \bar{s})) G_l(s_l) \quad (1)$$

$$D(\bar{s}) = D_{total} - G(\bar{s}) \quad (2)$$

Let $R_l(s_l)$ denote the total number of bits consumed by segment l when strategy s_l is applied, then the overall rate usage $R(\bar{s})$ can be calculated as

$$R(\bar{s}) = \sum_{l=1}^{L} R_l(s_l) \quad (3)$$

The cross-layer JSCC optimization problem becomes

$$\arg\min_{\bar{s}} D(\bar{s}) \text{ s.t. } R(\bar{s}) \le R_{max} \quad (4)$$

which minimizes the expected end-to-end distortion given a rate constraint. R_{max} is the total available bit rate, which equals to the bandwidth times the available transmission time. The dual problem is

$$\arg\min_{\bar{s}} R(\bar{s}) \text{ s.t. } D(\bar{s}) \le D_{max} \quad (5)$$

where D_{max} indicates the maximum tolerable distortion. If bandwidth is fixed, the equals to minimizing transmission time.

To find optimal solutions, dynamic programming techniques can be applied. However, if dynamic programming is used, the complexity can increase exponentially with L and the size of the strategy set. In Section 3, we propose a scalable cross-layer rate allocation algorithm with low complexity to approximate the optimal solutions, which works for both (4) and (5) .

3. SCRA: SCALABLE CROSS-LAYER RATE ALLOCATION ALGORITHM

For each segment l, let w_l denote the corresponding bit weighting factor, n_l the total number of bits without redundancy, $E_l(s_l)$ the average number of useless bits if one bit is corrupted, and $P_e(s_l)$ the residual BER after channel decoding, then we can have

$$G_l(s_l) = w_l n_l (1 - P_e(s_l) E_l(s_l)) \quad (6)$$

$$R_l(s_l) = n_l R_l^s(s_l^s) R_l^f(s_l^f) R_l^r(s_l^r) R_l^p(s_l^p) \quad (7)$$

Algorithm 1 feasible strategy set search

Let S_l be the set of all available schemes for segment l;
for $(\forall s \in S_l)$ **do**
  Calculate $G_l(s)$ and $R_l(s)$;
end for
Sort the elements s's of S_l according to $R_l(s)$;
Delete all the s's from S_l if there exists an s' such that $G_l(s') > G_l(s)$ and $R_l(s') \le R_l(s)$;
Perform convex hull analysis on S_l such that the elements of S_l satisfy the inequalities (11) and (12);
Return S_l;

where $R_l^s(s_l^s)$, $R_l^f(s_l^f)$, $R_l^r(s_l^r)$, and $R_l^p(s_l^p)$ denote the source error resilient coding redundancy ratio, the inverse of SF code rate, the inverse of RS code rate, and the inverse of PE code rate applied on segment l. The basic idea of the SCRA algorithm is to allocate the available rate step by step, at each step, it assigns some available bits to certain segment which can best utilize these bits. Before describing the algorithm, we first give some definitions.

For each segment l, let $\Delta G_l(s_l, s_l^{'})$ and $\Delta R_l(s_l, s_l^{'})$ be the *relative distortion reduction* and *relative rate consumption* by changing the schemes from s_l to $s_l^{'}$, that is:

$$\Delta G_l(s_l, s_l^{'}) = G_l(s_l^{'}) - G_l(s_l) \quad (8)$$

$$\Delta R_l(s_l, s_l') = R_l(s_l') - R_l(s_l) \quad (9)$$

and define the *normalized gain in rate* as

$$g_l(s_l, s_l') = \frac{\Delta G_l(s_l, s_l')}{\Delta R_l(s_l, s_l')} \quad (10)$$

Let $s^0, s^1, s^2, \ldots, s^{|S|} \in S$ be the set of all available schemes in the increasing order according to their rate consumption for a certain segment. Now we define the *feasible strategy set*, S_l, for each segment l, as the subset of S where the necessary and sufficient conditions that $s^i \in S$ belongs to S_l are

$$\min_{k<i} g_l(s^k, s^i) \quad > \quad 0 \quad (11)$$

$$\min_{k<i} g_l(s^k, s^i) \quad > \quad \max_{j>i} g_l(s^i, s^j) \quad (12)$$

In other words, all the feasible strategies should reside on the convex hull of the rate-distortion curve for that segment.

For each segment l, the feasible strategy set can be obtained using a conventional convex hull analysis presented in Alg. 1. This procedure is used to remove those schemes with less probability to be selected, reduce the scheme set size, and facilitate the strategy selection in the SCRA algorithm. After the feasible scheme sets for all segments have been obtained, the cross-layer rate allocation problem (4) can be approximated using the proposed SCRA algorithm presented in Alg. 2, which is a generalization of the marginal analysis algorithm in [6]. The algorithm first initializes the strategy for each segment to be s_l^0, which means no quality segment will be transmitted. At each iteration, it finds the segment l and strategy s_l which has the largest weighted D-R ratio, that is, largest $(1 - P_{drop}(l, \bar{s}))g_l(s_l, s_l')$, among all unmarked feasible strategies, then adjusts the number of bits consumed, marks s_l, and updates $P_{drop}(j, \bar{s})$ for all segments j's with $j > l$. After the constraint is satisfied, it then adjusts the length of the last quality segment to be transmitted to make the constraint to be tight. By making little

Algorithm 2 SCRA algorithm for cross-layer rate allocation

$R = 0$;
for $(1 \le l \le L)$ **do**
  Find S_l using Alg. 1;
  Let $s_l = s_l^0$ and mark s_l^0;
end for
while $(R < R_{max})$ **do**
  Find the segment l and the strategy s_l' such that $(1 - P_{drop}(l, \bar{s}))g_l(s_l, s_l')$ is maximized among all unmarked strategies of all segments;
  $R = R + \Delta R_l(s_l, s_l')$;
  Let $s_l = s_l'$ and mark s_l';
  Update $P_{drop}(j, \bar{s})$ with $j > l$;
end while
if $(R > R_{max})$ **then**
  Let l be the last segment with $s_l \neq s_l^0$, reduce the its length from n_l to n_l' such that $R - R_{max} = (n_l - n_l')R_l(s_l)$;
end if
Return $\bar{s}$ and $\{n_l\}$.

modification, the algorithm can also be used to approximate the solution for problem (5).

The complexity of the SCRA algorithm is analyzed below. Assume the total number of segments is L, the total number available strategies for each segment is N, and the number of marked strategies is M. Let the time needed to calculate $G_l(s_l)$ and $R_l(s_l)$ be T_g. By using quick sort, the complexity of the feasible strategy set search for each segment becomes $O(NT_g + Nlog_2N + 2N)$, and the overall complexity becomes $O(LNT_g + LNlog_2N + 2NL + LMlog_2L)$. If $L = 10$, $N = 100$, $M = 50$, the bound becomes $1000T_g + 10000$. The complexity can be further reduced if the segments have similar profiles.

4. SIMULATION RESULTS

In our simulations, a MIMO-OFDM system with 2 transmit and 2 receive antennas was used at PHY layer of wireless link. The variable rate SF code applied at DLL layer of wireless link is an extension of Blum's 4-state space time trellis code with QPSK modulation [2], and the BER vs. SNR curves were obtained via computer simulations, where a two-ray channel model was used with a delay of $20\mu s$. The transmitted image was a gray 512×512 *lenna* with 8 bits per pixel. The image was first compressed into 10 equal size quality segments with compression ratio 0.5bpp, and each quality segment was partitioned into 5 IP packets. Before transmission of each IP packet, header information was added, which introduced some extra overhead. For each segment, the adjustable parameters included the RS code rates, the PE code rates, the SF code rates, and the source error resilient coding schemes.

The simulation results for minimizing distortion under fixed rate constraint is presented in Fig. 2, where 0.5bpp means that the total number of available bits for transmission was $0.5 \times 512 \times 512$. Four situations were considered, and the SCRA algorithm was used in all cases. In the four situations, *UDP-lite* denotes that UDP-lite was used by the wireless link, *UDP* denotes that only traditional UDP protocol was used, and *PDR* denotes the packet dropping rate caused by congestion. From the results, we can see that at least 1.5dB gain can be achieved in PSNR by the *UDP-lite* over *UDP*. When the PDR becomes 10%, UDP-lite still has

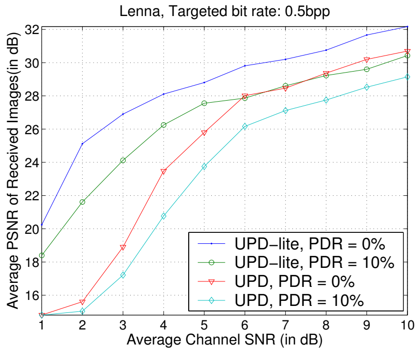

Fig. 2. Performance comparison under rate constraint

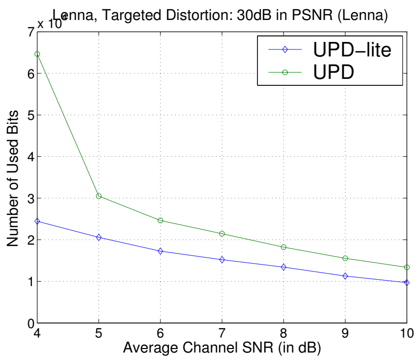

Fig. 3. Performance comparison under distortion constraint

better performance than UDP over almost all SNR values (1-8), especially when the SNR is low. This performance improvement comes from the usefulness of the erroneous packets to improve the reconstructed image quality and the advantage of using bit-level interleaving and channel coding at APL layer.

The simulation results for minimizing the transmission time under a given end-to-end PSNR constraint of reconstructed image is presented in Fig. 3. In the simulations, we assumed that the bandwidth does not change during transmission of each image, so minimizing transmission time equals to minimizing total number of bits needed. The cross-layer optimization is performed using the modified SCRA algorithm for problem (5). In the simulation we set the targeted PSNR 30 dB. From the results we can see that at least 30% can be saved in bits (and, consequently, in transmission time) by applying UDP-lite over traditional UDP.

We also compared the performance of the system using the SCRA algorithm with the system proposed in [7]. To achieve fair comparison, we modified our system such that the image was not transmitted on a packet network, but on a wireless channel without packet dropping. The differences between the two systems lie in that in [7], a fixed-rate space time block code was used, and the source coding standard is SPIHT, which has similar compression efficiency as JPEG 2000 but is less error resilient. Another difference is that the authors of [7] assumed that any erroneous bit in the received data will cause the following bit to become useless, so their objective was to maximize the number of correctly received bits before any error occurs. The simulation results in Fig. 4 show that our system outperforms the system in [7] in all situations, especially when the SNR is low.

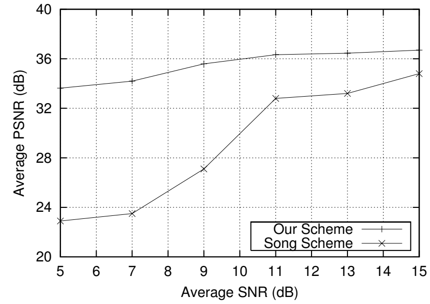

Fig. 4. Performance comparison of the proposed system with [7]

During our simulations, we also recorded the execution time needed for the SCRA algorithm. Using an IBM Thinkpad with CPU Pentium III 1.13GHz, RAM 256M, running Linux 2.2.14, the total time needed to perform SCRA on lenna image was 10ms with 10 segments and parameter set sizes of 256 for each segment.

5. CONCLUSIONS

In this paper, we presented an efficient real-time JPEG 2000 image transmission system over heterogeneous wireless networks, taking advantage of cross-layer optimization and UDP-lite protocol to utilize the limited bandwidth more efficiently. To manage the large number of adjustable parameters across network layers, a scalable cross-layer rate allocation (SCRA) algorithm was proposed to perform the cross-layer rate allocation with low complexity. The simulation results showed the advantages of exploiting cross-layer optimization and UDP-lite, where for *lenna* image, at least 1.5dB gain in PSNR can be obtained for a fixed transmission rate of 0.5bpp, or at least 30% transmission time can be saved under a targeted PSNR of 30dB over the systems without exploiting UDP-lite. The efficiency of the proposed system was also shown by comparing it with a similar system, without considering packet dropping and with a fixed transmission rate of 0.5bpp, 6dB gain in PSNR can be obtained when the channel SNR is 10dB. The simulation also confirms the low complexity of the SCRA algorithm.

6. REFERENCES

[1] L-A. Larzon, M. Degermark, and S. Pink, "The udp-lite protocol," IETF Network Working Group, Dec. 2002.

[2] W. Su, Z. Safar, M. Olfat, and K. J. R. Liu, "Full-diversity space-frequency codes for mimo-ofdm systems," in *IEEE ISIT*, June 2003.

[3] ISO/IEC, "ISO/IEC 15444-1:2000: Information technology — JPEG 2000 image coding system — part 1: core coding system," 2000.

[4] Luigi Rizzo, "Effective erasure codes for reliable computer communication protocols," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 2, pp. 24–36, 1997.

[5] S. Lin and D.J. Costello, *Error Control Coding*, Prentice-Hall, 1983.

[6] B. Fox, "Discrete optimization via marginal analysis," *Management Science*, vol. 13, no. 3, pp. 210–216, Nov. 1966.

[7] J. Song and K. J. R. Liu, "Robust progressive image transmission over ofdm systems using space-time block code," *IEEE Transactions on Multimedia*, vol. 4, no. 3, September 2002.