8-2011

A Cross-Layer Design for the Performance Improvement of Real-Time Video Transmission of Secondary Users Over Cognitive Radio Networks

Haiyan Luo
University of Nebraska-Lincoln, haiyan.luo@huskers.unl.edu

Song Ci
University of Nebraska-Lincoln, sci2@unl.edu

Dalei Wu
University of Nebraska-Lincoln, dwu@huskers.unl.edu

A Cross-Layer Design for the Performance Improvement of Real-Time Video Transmission of Secondary Users Over Cognitive Radio Networks

Haiyan Luo, *Student Member, IEEE,* Song Ci, *Senior Member, IEEE,* Dalei Wu, *Student Member, IEEE*

***Abstract*—Cognitive radio (CR) has been proposed as a promising solution to improve connectivity, self-adaptability, and efficiency of spectrum usage. When used in video applications, user-perceived video quality experienced by secondary users is a very important performance metric to evaluate the effectiveness of CR technologies. However, most of the current research only considers spectrum utilization and effectiveness at medium access control (MAC) and physical layers, ignoring the system performance of the upper layers. Therefore, in this paper, we aim to improve the user experience of secondary users for wireless video services over CR networks. We propose a quality-driven cross-layer optimized system to maximize the expected user-perceived video quality at the receiver end under the constraint of packet delay bound. By formulating network functions such as encoder behavior, cognitive MAC scheduling, transmission, as well as modulation and coding into a distortion-delay optimization framework, important system parameters residing in different network layers are jointly optimized in a systematic way to achieve the best user-perceived video quality for secondary users in CR networks. Furthermore, the proposed problem is formulated into a MIN-MAX problem, and solved by using dynamic programming. The performance enhancement of the proposed system is evaluated through extensive experiments based on H.264/AVC.**

***Index Terms*—Cognitive radio (CR), cross-layer, dynamic programming, MIN-MAX problem, primary users, secondary users, video distortion, wireless video transmission.**

I. Introduction

WITH THE FAST development of wireless communication technologies, the limited unlicensed spectrum bands can no longer meet the increasing requirement. Frequency spectrum has become the scarcest resource in the next-generation wireless networks. However, statistics indicate that a large amount of the licensed spectrum bands are under-utilized, due to the spectrum strategy of static allocation and centralized management. Research also shows that the utilization of both licensed and unlicensed frequency bands changes with time and location. Thus, efficient methods are necessary for spectrum sharing among different systems, applications, and services in a dynamic wireless environment [1]. Therefore, "cognitive radio" (CR) has emerged as a new design paradigm for next-generation wireless networks, aiming to increase the utilization of radio spectrum for both licensed and unlicensed bands.

Generally speaking, CR is an intelligent wireless communication system that is aware of its surrounding environment and uses the methodology of understanding-by-building to learn from the environment. Then, its internal states are adapted to respond to statistical variations of the incoming radio frequency stimuli by changing certain operating parameters in real-time. Here, there are two primary objectives: 1) to provide highly reliable communications whenever and wherever needed, and 2) to achieve efficient utilization of the radio spectrum [2]. Fig. 1 illustrates a typical scenario of wireless video transmission over CR networks, where primary users and secondary users share a block of frequency spectrum. In CR networks, only primary users are authorized to use the radio spectrum. Thus, secondary users have to search the idle channels to use at the beginning of every slot by performing channel sensing. Based on the sensing outcomes, secondary users will decide whether or not to access the sensed channels. It has been reported that some frequency bands in the radio spectrum are largely unused, while some are heavily used. In particular, while a frequency band is assigned to a primary wireless system/service at a particular time and location, the same frequency band is unused by this wireless system/service in other times and locations. This results in spectrum holes (known as spectrum opportunities) [3]. Therefore, by allowing secondary users to utilize these spectrum holes, spectrum utilization can be improved substantially.

Indeed, improving system performance of secondary users is a vital factor that hinges on the success of CR technologies for wireless video applications. Actually, if upper layer system performance is not well-considered, the video quality degradation perceived by secondary users can largely impede the successful deployment of CR technologies. Also, user experience is by far the most important performance metric for end users in wireless video transmission. Therefore, it is critical to improve the user-perceived video quality at the receiver side for secondary users in CR networks. Although a

Manuscript received May 4, 2010; revised October 16, 2010; accepted November 22, 2010. Date of publication March 17, 2011; date of current version August 3, 2011. This work was supported in part by the NSF, under Grants CCF-0830493 and CNS-0707944. This paper was recommended by Associate Editor Y. Qian.

H. Luo and D. Wu are with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588 USA (e-mail: haiyan.luo@huskers.unl.edu; dwu@huskers.unl.edu).

S. Ci is with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588 USA, and also with the HPNL Laboratory, IOA, Chinese Academy of Sciences, Beijing 100190, China (e-mail: sci@engr.unl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2129810

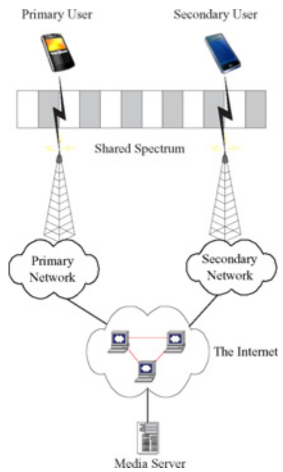


Fig. 1. Wireless video transmission over CR networks, where primary users and secondary users access the network through shared radio frequency band.

lot of research activities have been conducted in CR networks, most of them only consider sensing effectiveness and spectrum utilization at the system design level. Other design metrics of the upper layers, such as the user-perceived video quality, have been mostly ignored.

In this paper, to achieve the best user-perceived video quality at the receiver side for secondary users of real-time wireless video transmission over CR networks, we propose a quality-driven cross-layer system for joint optimization of system parameters residing in the entire network protocol stack. In the proposed system, time variations of primary network usage and wireless channels are modeled, based on which, the encoder behavior, cognitive MAC scheduling, transmission, and modulation and coding are jointly optimized for secondary users in a systematic way under a distortion-delay framework for the best video quality perceived by secondary users. The proposed problem is formulated as a MIN-MAX problem and solved by using dynamic programming [4], [5]. Furthermore, the experiments prove that the proposed joint optimization system can greatly improve user experience at the receiver side for real-time wireless video services over CR networks. To the best of our knowledge, this is the first cross-layer design work to improve the performance of CR networks for real-time wireless video transmission from the perspective of application services. Another contribution of this paper is that we quantify and improve the video performance for secondary users so that the deployability of the CR technologies can be increased.

The remainder of this paper is organized as follows. We first introduce the peer work related to this topic in Section II. Then, in Section III, we present the proposed system model of quality-driven cross-layer optimization. In Section IV, we discuss the problem formulation, followed by the optimal solution in Section V. The details of experimental study and analysis are described in Section VI. Finally, Section VII concludes this paper.

II. Related Work

Cross-layer design over CR networks was briefly discussed in [6] and [7]. In the proposed frameworks, each node in the network can sense and learn from the wireless environment and then respond to environment changes by adapting system parameters. However, how to integrate the different layers was not discussed. Also, there was no experimental results to verify the proposed frameworks.

In [8] and [9], a rake optimized power-aware scheduling architecture was proposed for mobile ad hoc networks to deal with the utilization of CR for dynamic channel allocation among the requesting applications while limiting the average power transmitted in each sub-band. In that work, the cross-layer interaction between medium access control (MAC) and physical (PHY) layers of CR networks was considered. The joint power control and link scheduling can reduce adjacent channel and multi-access interference.

Su *et al.* [10] have proposed an opportunistic multichannel MAC protocol, in which the spectrum sensing at the PHY layer and the packet scheduling at the MAC layer are integrated for wireless ad hoc networks. The proposed MAC protocol enables secondary users to identify and utilize the leftover frequency spectrum to reduce the interference level. Two different policies on channel sensing, including random sensing policy and negotiation-based sensing policy, have been proposed to detect the availability of unused licensed channels.

In [11], a framework called cognitive resource manager (CMR) was proposed to optimize the network protocol stack as a whole. The exchange of network information between CMRs can avoid harmful interactions arising from local optimization methods. In addition, the proposed framework can adapt MAC and link parameters to choose the best possible settings for the applications running on top. However, no implementation details and experimental analysis were discussed.

Overall, most current research efforts on cross-layer design over CR networks only focus on the joint consideration of MAC layer and PHY layer. The design objectives are limited due to the fact that only sensing effectiveness and spectrum utilization are used as the design criteria, while the performance at the upper layers has been largely ignored.

III. System Model

In this paper, the proposed quality-driven cross-layer optimized system includes different modules, such as video encoder module, cognitive MAC module, modulation and coding module, cross-layer optimization module, as well as wireless video transmission module. These system modules actually represent different network functions residing in different network layers. For example, the video encoder resides in the application layer. The cross-layer optimization module is the core of the proposed system. The cognitive MAC module resides in the MAC layer, while the modulation and coding module is in the PHY layer. All these modules exist in the same network node. As shown in Fig. 2, the cross-layer optimization module is able to communicate with other system modules to adjust the network functions, by selecting the optimal system parameters within a distortion-delay optimization framework. In this way, the major network functions are jointly optimized to achieve the best user-perceived video quality over CR networks under the current network conditions.

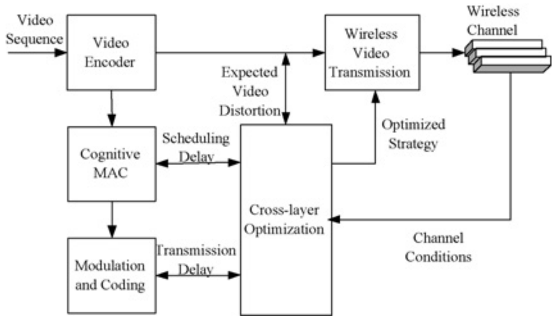


Fig. 2. Proposed quality-driven cross-layer optimized system model for real-time wireless video transmission over CR networks.

In the proposed system, the expected video distortion calculated at the video encoder of the application layer is adopted as the objective function, which can be represented by encoding parameters (such as quantization step size q) and packet loss rate ρ. Furthermore, the expected packet delay, which is jointly represented by system parameters that affect encoder behavior, MAC scheduling, transmission, and modulation and channel coding, is used as the design constraint. Therefore, with the feedback information from the network such as round trip time, queue length, and packet loss rate, the distortion-delay optimization module can choose the optimal set of parameters through the proposed cross-layer optimized system to achieve the best user-perceived video quality.

A. Video Distortion

One of the design challenges is to estimate the end-to-end video distortion at the video encoder to be used as the objective function in the proposed design. Video encoder situates at the application layer. Without losing generality, we consider H.264 video codec in this paper. In H.264 codec, each video frame is represented by block-shaped units of the associated luminance and chrominance samples (16 × 16 pixel region) called macroblocks (MBs). Furthermore, MBs can be both intra-coded or inter-coded from samples of previous video frames [12]. Intra-coding is performed in the spatial domain by referring to neighboring samples of previously-coded blocks which are on the left and/or above the block to be predicted. Meanwhile, inter-coding is performed with temporal prediction from samples of previous video frames.

To estimate the end-to-end video distortion accurately, we need to consider all possible factors, which include source coding, error propagation, channel coding, and others. Many research efforts can be found in the literature on distortion estimation for hybrid motion-compensated video coding and transmission over lossy channels [13]–[16]. For real-time source coding, the estimated distortion caused by quantization, packet loss, and error concealment at the encoder can be calculated by using the "recursive optimal per-pixel estimate" (ROPE) method [13], providing an accurate optimization metric to the proposed system based on video quality. For the source coding parameter, we consider quantization step size (QP) q in this paper.

According to the H.264 standard, one packet can be set to be one row of MBs, which is also called one slice [12]. Therefore, slice and packet are two interchangeable concepts in this paper. Given the dependencies introduced by error concealment scheme, the expected distortion of packet x of video frame n can be calculated at the encoder by using ROPE method as

$$\begin{aligned} E[D_{n,x}] &= (1-\rho_x)E[D^r_{n,x}] \\ &+ \rho_x(1-\rho_{x-1})E[D^{lr}_{n,x}] + \rho_x\rho_{x-1}E[D^{ll}_{n,x}] \end{aligned} \tag{1}$$

where ρ_x is the loss probability of packet x with consideration of packet delay bound T_n^{max}. $E[D^r_{n,x}]$ is the expected distortion of packet x when it is successfully received. Furthermore, depending on whether packet $(x-1)$ is received or lost, $E[D^{lr}_{n,x}]$ and $E[D^{ll}_{n,x}]$ are the corresponding expected distortion after concealment when packet x is lost. Therefore, the expected distortion of the whole video frame n can be represented as

$$E[D_n] = \sum_{x=1}^{X_n} E[D_{n,x}] \tag{2}$$

where X_n is the total number of packets in the video frame n. Thus, the expected end-to-end video distortion is accurately calculated by ROPE under instantaneous network conditions, which becomes the objective function in the proposed optimized system. For a given video packet x, the expected packet distortion only depends on packet error rate ρ_x and QP q. Considering the fact that the individual contribution of each path is continuously updated, this parameter is updated after each packet is encoded. In addition, the prediction and calculation of packet loss rate ρ_x will be discussed in Section III-D. The readers can also refer to [4] and [13] for detailed information regarding the calculation of the expected video distortion.

B. Channel Model

In this paper, we assume that wireless channels are frequency flat, remaining time-invariant during a packet, but may vary from packet to packet. The channel quality is captured by the received signal-to-noise ratio (SNR) ξ. We adopt the Rayleigh channel model to describe ξ statistically. Therefore, the received SNR per packet is a random variable with a probability density function (pdf) as follows:

$$p(\xi) = \frac{1}{\bar{\xi}}\exp(-\frac{\xi}{\bar{\xi}}), \qquad \xi \geq 0 \tag{3}$$

where $\bar{\xi} := E\{\xi\}$ is the average received SNR. We also assume that the receiver has perfect channel side information and hence knows the instantaneous values of channel state information (CSI). However, the channel quality is fed back to the transmitter constantly so that the receiver can dynamically adapt the varying channel status for better system performance. The feedback frame size is very small and its transmission takes a high priority. Thus, the delay effect on the adaptation results can be ignored [5], [17], [18].

Moreover, we employ a cognitive channel model in which secondary users will try to transmit data when primary users are in presence. Secondary users will first perform channel sensing to detect the activity of primary users, then decide

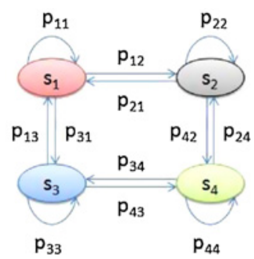


Fig. 3. Proposed state transition model for CR channel.

whether to transmit the data immediately or wait for the next available time slot depending on the detection result.

Define the following four states of a given channel for primary usage.

1) State s_1: Channel is idle, detected as idle (correct detection).
2) State s_2: Channel is idle, detected as busy (false alarm).
3) State s_3: Channel is busy, detected as idle (missed detection).
4) State s_4: Channel is busy, detected as busy (correct detection).

Based on this definition, we depict the state transition model for CR transmission as shown in Fig. 3. In this figure, the state transition model is completely described by its stationary distribution of each channel state i, and the state transition probability from state s_i to state s_j at the beginning of each time slot, which is denoted as p_{ij} $(1 \leq i, j \leq 4)$. Given the knowledge of channel fading, state transition probabilities, and primary user usage, the channel transition matrix $\mathbf{R}$ can be written as

$$\mathbf{R} = \{p_{ij}\} \quad (1 \leq i \leq 4, 1 \leq j \leq 4) \tag{4}$$

where

$$p_{ij} \geq 0, \forall i, j$$
$$\sum_{j=1}^{4} p_{ij} = 1, \forall i.$$

Next, we will derive the transition matrix $\mathbf{R}$. In this paper, we adopt noisy observations y_i under the Neyman-Pearson formulation to detect correlated random signals [19]–[21]. Therefore, channel sensing can be formulated as a hypothesis testing problem between the noise w_i and the signal s_i in noise as follows:

$$\begin{aligned} H_0: &\quad y_i = w_i, \quad i = 1, ..., S_{tot} \\ H_1: &\quad y_i = s_i + w_i \quad i = 1, ..., S_{tot} \end{aligned} \tag{5}$$

where S_{tot} is the total symbols within a duration that is allocated to sense the channel. Received noise $\{w_i\}$ is a zero-mean, complex Gaussian random variable with variance σ_w^2 for all i, denoted as $\{w_i\} \sim S_{tot}(0, \sigma_w^2)$. $\{s_i\}$ is the sum of the active primary users' faded signals arriving at the secondary receiver, which has a circularly symmetric complex Gaussian distribution with zero-mean and variance σ_s. Both $\{s_i\}$ and $\{w_i\}$ are assumed to be independent and identically distributed [21].

Therefore, the optimal Neyman-Pearson detector of the above detection problem is given by

$$Y = \frac{1}{S_{tot}} \sum_{i=1}^{S_{tot}} |y_i^2| \gtrless_{H1}^{H_0} t_d \tag{6}$$

where t_d is the detection threshold. Thus, test statistic Y is chi-square distributed with $2S_{tot}$ degrees of freedom. Therefore, the probabilities of detection p_d and false alarm p_f can be represented as follows:

$$p_d = P_r(Y > t_d|H_1) = 1 - P(\frac{S_{tot}t_d}{\sigma_w^2 + \sigma_s^2}, S_{tot}) \tag{7}$$

$$p_f = P_r(Y > t_d|H_0) = 1 - P(\frac{S_{tot}t_d}{\sigma_w^2}, S_{tot}) \tag{8}$$

where $P(x, z)$ denotes the regularized lower gamma function. Denote $\Gamma(z)$ and $\gamma(x, z)$ the Gamma function and the lower incomplete gamma function, respectively, then $P(x, z)$ can be represented as

$$P(x, z) = \frac{\gamma(x, z)}{\Gamma(z)}. \tag{9}$$

Therefore, the probability of channel being idle and detected as idle can be calculated as follows:

$$p_{11} = (1 - p_f)(1 - \zeta_b) \tag{10}$$

where ζ_b is the prior probability of channel being busy. Equation (10) indicates that the transition probability only depends on the current state, regardless of the original state. This also holds true for the other state transition probabilities. Denote

$$\begin{aligned} &p_{i1} = p_1; \quad p_{i2} = p_2 \\ &p_{i3} = p_3; \quad p_{i4} = p_4 \\ &\text{where} \quad 1 \leq i \leq 4. \end{aligned} \tag{11}$$

Then, the transition probabilities can be represented as

$$\begin{aligned} p_1 &= (1 - p_f)(1 - \zeta_b) \\ p_2 &= p_f(1 - \zeta_b) \\ p_3 &= (1 - p_d)\zeta_b \\ p_4 &= p_d\zeta_b. \end{aligned} \tag{12}$$

Hence, the 4×4 transition matrix $\mathbf{R}$ can be derived as

$$\mathbf{R} = \begin{pmatrix} p_1, p_2, p_3, p_4 \\ p_1, p_2, p_3, p_4 \\ p_1, p_2, p_3, p_4 \\ p_1, p_2, p_3, p_4 \end{pmatrix}. \tag{13}$$

C. MAC Scheduling Delay

To formulate the MAC frame scheduling delay for secondary users in CR networks, we first denote T_0 the duration of a time slot, and t_s the corresponding channel sensing time allocated for each time slot, as shown in Fig. 4.

In the proposed channel model shown in Fig. 3, if the current channel is in state s_2 or s_4, the MAC frame has to wait for the next time slot. When new time slot arrives, it

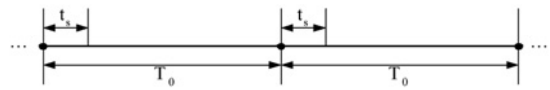


Fig. 4. Illustration of time slot duration and channel sensing time over CR networks.

has to wait again if the channel is still in state s_2 or s_4. This process repeats until a time slot becomes available, or until a maximum waiting threshold in terms of the number of time slots is reached, denoted as N_s^{max}. If this maximum threshold is reached, the waiting packet has to be dropped from the sending queue. We call this truncated MAC scheduling. Furthermore, to ensure real-time video transmission, every video packet has to meet a delay bound. Therefore, all the video packets of the same video frame have the same delay bound. We denote T_n^{max} the delay bound of the video frame n. Then, N_s^{max} can be calculated as

$$N_s^{max} = \left\lfloor \frac{T_n^{max}}{T_0} \right\rfloor. \tag{14}$$

Note that forward error control coding and automatic repeat request (ARQ) are the two major error resilient approaches used by video encoder. However, ARQ is not always feasible for real-time video transmission, due to the excessive delay caused by retransmissions [22], [23]. Therefore, in this paper, we do not consider ARQ. Instead, we will optimize system parameters in a holistic way to improve the overall system performance for secondary users.

As shown in Fig. 3, only when channel is in state s_1, can it be effectively used to transmit data for secondary users. When the channel is in state s_2, s_3, or s_4, it is not available for secondary users, or it is not detected by the secondary users as available. Therefore, the probability p_w that a given MAC frame has to wait for the next time slot can be expressed as

$$p_w = 1 - (1 - p_f)(1 - \zeta_b). \tag{15}$$

Thus, assuming the availability of time slots are independent, the average scheduling time at hop h can be represented as

$$\begin{aligned} t_{sched}^h &= (p_w + p_w^2 + \ldots + p_w^{N_s^{max}}) * T_0 \\ &= \left(\frac{p_w - p_w^{N_s^{max}}}{1 - p_w} \right) * T_0. \end{aligned} \tag{16}$$

D. Transmission Delay

In this paper, adaptive modulation and coding (AMC) technique at the PHY layer is adopted on a packet-by-packet basis to enhance the throughput. With AMC, the optimal combination of different modulation constellations and different rates of error-control codes is selected based on the time-varying channel quality. For example, in good channel conditions, AMC schemes with larger constellation sizes and higher channel coding rate will guarantee the required packet error rate for quality of service provisioning [24], [25]. Usually, bit error rate (BER) $\epsilon(\xi)$ can be calculated from the approximated expression as follows:

$$\epsilon(\xi) = a_m e^{-b_m \xi} \tag{17}$$

where coefficients a_m and b_m can be obtained by fitting (17) to the exact BER, as shown in Table I [18]. Therefore, the frame error rate $\varrho(\xi)$ can be expressed as

$$\varrho(\xi) = 1 - (1 - \epsilon(\xi))^{\ell} \tag{18}$$

where ℓ is the MAC frame size, and ξ is the received SNR.

Assume that the xth packet of video frame n is fragmented into Z MAC frames, and these MAC frames are transmitted along a path with H hops. Thus, the packet loss rate for a given packet x can be expressed as

$$\rho_x = 1 - \prod_{z=1}^{Z} (1 - \varrho_x^z(\xi)) \tag{19}$$

where $\varrho_x^z(\xi)$ is the frame error rate of MAC frame z of packet x.

Denote BW_h the bandwidth (symbols/s) of hop h, and $\ell_x^z(m, h)$ the bit number of the zth PHY frame of packet x at hop h. Therefore, the transmission delay can be represented as

$$t_{trans}^h = \left\lceil \frac{\frac{\ell_x^z(m,h)}{r_{m,h} \times BW_h}}{T_0 - t_s} \right\rceil * T_0 \tag{20}$$

where $r_{m,h}$ is the rate (bits/symbol) of AMC mode m at hop h as shown in Table I.

IV. Problem Formulation

To achieve the best video quality at the receiver side, the expected end-to-end video distortion under the constraint of video packet delay should be minimized. With the proposed distortion-delay framework, the source coding, MAC scheduling, transmission, and modulation and coding are jointly optimized in a cross-layer fashion.

In wireless video transmission, all packets of a given frame f_n is constrained by a frame delay bound $T_n^{\max}$. Therefore, all packets of the video frame f_n have the same delay constraint $T_n^{\max}$. Denote $\mathbf{Q}$ as all possible operating points of source coding parameter (such as quantization step size $q_{n,x}$) of packet x of frame n, $\mathbf{M}$ as all possible modulation and channel coding schemes $m_{n,x}$. Thus, the proposed problem can be formulated as

$$\begin{aligned} &\min_{\{q_{n,x} \in \mathbf{Q}, m_{n,x} \in \mathbf{M}\}} \sum_{n=1}^{N} \sum_{x=1}^{X_n} E[D_{n,x}] \\ &\quad s.t. : \max_{1 \le x \le X_n} t_{n,x} \le T_n^{max}, \quad \forall x \end{aligned} \tag{21}$$

where N is the total number of video frames of the given video sequence, and X_n is the total number of packets generated from the nth video frame. In addition

$$t_{n,x} = \sum_{z=1}^{Z} \sum_{h=1}^{H} (t_{sched}^{z,h} + t_{trans}^{z,h}) \tag{22}$$

where $t_{sched}^{z,h}$ and $t_{trans}^{z,h}$ have already been derived by using (16) and (20), respectively. Z and H are the MAC frame number and the hop number, respectively. In other words, the xth packet of video frame n is fragmented into Z MAC frames, and these MAC frames are transmitted on a path with H hops.

TABLE I
AMC SCHEMES AT THE PHY LAYER

AMC mode (m)	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
Modulation scheme	BPSK	QPSK	QPSK	16-QAM	16-QAM	64-QAM
Coding rate (c_m)	1/2	1/2	3/4	9/16	3/4	3/4
r_m (bits/symbol)	0.50	1.00	1.50	2.25	3.00	4.50
a_m	1.1369	0.3351	0.2197	0.2081	0.1936	0.1887
b_m	7.5556	3.2543	1.5244	0.6250	0.3484	0.0871

Therefore, the proposed problem has been formulated into a MIN-MAX problem [26].

V. PROBLEM SOLUTION

Denote α as the index of video packet x over the entire video clip. Thus, the parameter vector of packet x of video frame n can be represented as

$$\begin{aligned} \mathcal{V}_\alpha \quad &:= [q_{n,x}, m_{n,x}] \\ \text{where} \quad & q_{n,x} \in \mathbf{Q}, \quad m_{n,x} \in \mathbf{M} \\ \text{and} \quad & 1 \leq \alpha \leq N \times X_n \end{aligned} \tag{23}$$

where $q_{n,x}$ and $m_{n,x}$ are the quantization step size, the AMC mode of packet x of the nth video frame, respectively. In addition, $\mathbf{Q}$ and $\mathbf{M}$ are the sets of all the possible values of $q_{n,x}$ and $m_{n,x}$, respectively.

To solve the MIN-MAX problem as shown in (21), we first convert it into an unconstrained optimization problem. According to the formulation (21), any parameter vector $\mathcal{V}_\alpha$ resulting in the expected packet delay greater than the constraint T_n^{max} cannot be the optimal parameter vector $\mathcal{V}_\alpha^*$, defined as $\mathcal{V}_\alpha^* := [q_{n,x}^*, m_{n,x}^*]$. Therefore, the objective function can be re-written as

$$E[D_{n,x}] = \begin{cases} \infty : & t_{n,x} > T_n^{max} \\ E[D_{n,x}] : & t_{n,x} \leq T_n^{max} \end{cases} \tag{24}$$

where the average distortion of a packet with expected delay greater than the delay bound T_n^{max} is set to infinity, meaning that the corresponding parameter vector of the possible solution will not satisfy the packet delay bound T_n^{max}. In this way, the minimum distortion problem with delay constraint is transformed into an unconstrained optimization problem.

Note that most modern source codecs such as H.264 [12] adopt error concealment strategies to improve the visual quality, which also introduces dependencies among slices/packets. Therefore, if the error concealment algorithm uses the motion vector of the previous slice to recover the lost slice, it will cause the calculation of the expected distortion of the current slice to depend on its previous slice. As mentioned earlier, a packet is generated from one slice. Therefore, without losing generality, we assume that the current packet depends on its previous θ packets ($\theta \geq 0$) in error concealment. To solve the formulated optimization problem in (21), we define a cost function $f_{\mathbb{k}}(\mathcal{V}_{\mathbb{k}-\theta}, ..., \mathcal{V}_{\mathbb{k}})$ to represent the minimum average distortion up to and including the $\mathbb{k}$th packet, where $\mathcal{V}_{\mathbb{k}-\theta}, ..., \mathcal{V}_{\mathbb{k}}$ are decision vectors of the $(\mathbb{k}-\theta)$th to $\mathbb{k}$th packets. Denote τ the total packet number of the video sequence, where $\tau := N \times X_n$. Therefore, $f_\tau(\mathcal{V}_{\tau-\theta}, ..., \mathcal{V}_\tau)$ represents the minimum distortion incurred by all packets of the given video sequence. Thus, solving (21) is essentially to solve the equation as follows:

$$\min_{\{\mathcal{V}_{\tau-\theta}, ..., \mathcal{V}_\tau\}} f_\tau(\mathcal{V}_{\tau-\theta}, ..., \mathcal{V}_\tau). \tag{25}$$

Therefore, given the cost function $f_{\mathbb{k}-1}(\mathcal{V}_{\mathbb{k}-\theta-1}, ..., \mathcal{V}_{\mathbb{k}-1})$ and the $\theta+1$ decision vectors $\mathcal{V}_{\mathbb{k}-\theta-1}, ..., \mathcal{V}_{\mathbb{k}-1}$ for the $(\mathbb{k}-\theta-1)$th to the $(\mathbb{k}-1)$th packets, the selection of the next decision vector $\mathcal{V}_{\mathbb{k}}$ is independent of the selection of the previous decision vectors $\mathcal{V}_1, \mathcal{V}_2, ..., \mathcal{V}_{\mathbb{k}-\theta-2}$. This means that the cost function can be expressed recursively as

$$\begin{aligned} f_k(\mathcal{V}_{\mathbb{k}-\theta}, ..., \mathcal{V}_{\mathbb{k}}) &= \min_{\{\mathcal{V}_{\mathbb{k}-\theta-1}, ..., \mathcal{V}_{\mathbb{k}-1}\}} \{ f_{\mathbb{k}-1}(\mathcal{V}_{\mathbb{k}-\theta-1}, ..., \mathcal{V}_{\mathbb{k}-1}) \\ &+ \quad E[D_{\mathbb{k}}] \} \end{aligned} \tag{26}$$

which implies that the next step of the optimization process for the cost function is independent of its past steps, forming the foundation of dynamic programming.

Essentially, (26) can be further converted into and solved as a graph theory problem of finding the shortest path in a directed acyclic graph [27]. By using dynamic programming to solve this shortest path problem, the computational complexity of the algorithm is decreased to $\mathbf{O}(\tau \times |\mathcal{V}|^{\theta+1})$ (where $|\mathcal{V}|$ is the cardinality of $\mathcal{V}$), depending directly on the value of θ. For most cases, θ is a small number, so the computational complexity of the algorithm is effectively decreased, compared with the exponential computational complexity of exhaustive search algorithm [28].

VI. EXPERIMENTAL ANALYSIS

A. Experimental Environment

In this paper, video coding is performed by H.264/AVC JM 15.1 codec under Linux platform while the channel model is generated by a simulator. The video sequence *Foreman* is adopted for performance analysis. The first 100 frames of the quarter common intermediate format (176 × 144) video clip are coded at frame rate of 30 f/s, and each I frame is followed by nine P frames. Assume the whole packet/slice is lost if one of the MAC frames of the packet is lost, which is reasonable since usually intra-prediction is derived from the decoded samples of the same decoded slice. To avoid prediction error propagation, a 10% MB level intra-refreshment is used during the experiments. When a packet is lost during transmission, the temporal-replacement error concealment strategy will be used. The motion vector of a missing MB can be estimated as the median of motion vectors of the nearest three MBs in the preceding row. If that row is also lost, the estimated motion

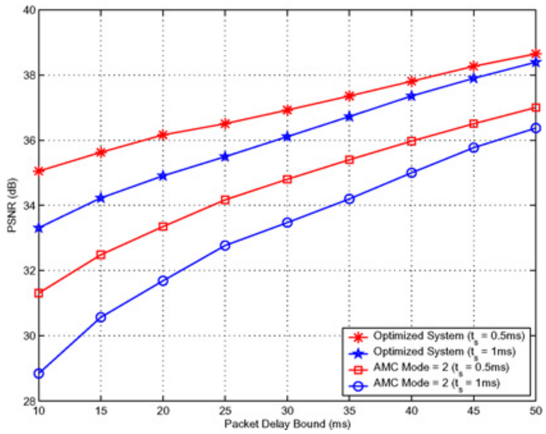


Fig. 5. Average PSNR comparison between the proposed quality-driven cross-layer optimized system and the existing system for real-time wireless video transmission over CR networks with various packet delay bounds T_n^{max}.

vector is set to zero. The pixels in the previous frame, pointed by the estimated motion vector, are used to replace the missing pixels in the current frame.

In the experiments, QP q and AMC mode m of each packet are considered as the parameters to be optimized. The possible values of QP are chosen from 1 to 50, while the available AMC schemes are 1 to 6 as shown in Table I. According to [21], the channel is assumed to be busy with an average probability of $\zeta_b = 0.1$. Also, since performance enhancement of secondary users in CR networks is the main focus of this paper, we set the detection threshold $t_d = 1.35$, so that false alarm probabilities are effectively decreased. Therefore, under this setting, the channel sensing is reliable, and the interference to primary users is minimal.

Given an average SNR $\bar{\xi}$, the instantaneous link quality ξ can be randomly produced from (3). In this paper, the link bandwidth is set to 100k symbols/s. Moreover, without losing generality, a single hop scenario is considered in the experiments to verify the performance of the proposed framework. Similar conclusions derived from the single hop scenarios may straightly apply to the multihop scenarios when the CSI of each hop is available.

Also, the delay bound is set in accordance with the frame rate, as adopted in the literature [29]–[31]. As the most important performance metric for video applications [32], [33], peak signal-to-noise ratio (PSNR) of the received video frames of secondary users is used as the performance metric to compare the proposed system with the existing system, which has fixed AMC schemes.

B. Performance Evaluation

The performance enhancement of the proposed system for secondary users under various packet delay bounds is verified in Fig. 5, where time slot duration T_0 is set to 5 ms, average SNR ξ is set to 15 dB, and channel sensing time t_s is set to 0.5 ms and 1 ms, respectively. From the figure, we can observe that by jointly optimizing the system parameters residing in different network layers under the proposed system, significant performance improvement can be achieved. Another observation is that as the delay bound becomes more and more stringent, the performance gain becomes higher and higher. This implies that the proposed system is especially suitable for real-time video transmission over CR networks with stringent delay bound. Furthermore, we can also observe that under the same network conditions, if the channel sensing time becomes longer, the performance gain actually becomes higher. This indicates that the proposed system might be useful for cognitive networks, due to the fact that when more time is allocated to perform channel sensing, MAC scheduling, and ARQ management, the negative impact on the overall system performance is minimized.

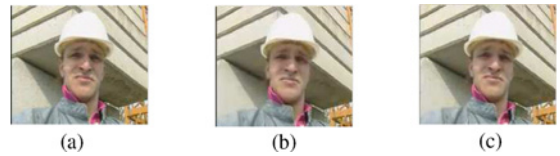


Fig. 6. Visual comparison between the proposed quality-driven cross-layer optimized system and the existing system for real-time wireless video transmission over CR networks. (a) Original frame. (b) Without optimization. (c) Optimized.

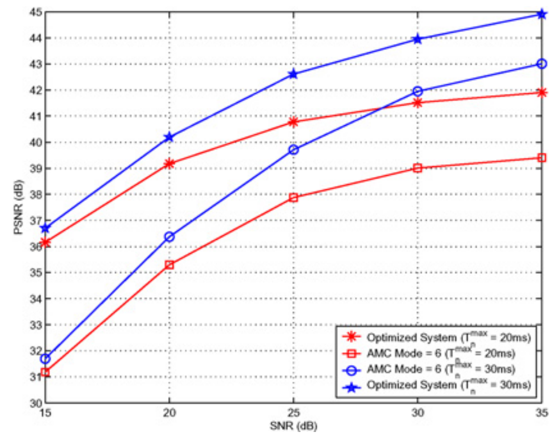


Fig. 7. Average PSNR comparison between the proposed quality-driven cross-layer optimized system and the existing system for real-time wireless video transmission over CR networks with various SNRs.

In Fig. 6, the visual comparison of one video frame (the 40th frame of the *Foreman* video clip) is presented with the packet delay bound being set to 30 ms and t_s 0.5 ms. Other environment settings remain the same with those of the above figure. Thus, it is observable that the user-perceived video quality has been greatly improved.

We also evaluate the relationship between the channel quality SNR ξ and the perceived video quality under the proposed system as shown in Fig. 7. In this experiment, time slot duration T_0 is set to 5 ms and the channel sensing time t_s is 0.5 ms. The packet delay bound is set to 20 ms and 30 ms, respectively. From the figure, we can observe that the proposed system can significantly improve PSNR performance, especially when the channel quality is not good or/and the delay bound is more stringent.

Furthermore, the impact of time slot duration T_0 on secondary users is shown in Fig. 8. In this case, the channel sensing time t_s is set to 0.5 ms and average SNR ξ is 15 dB. The packet delay bound T_n^{max} is set to 20 ms and 30 ms,

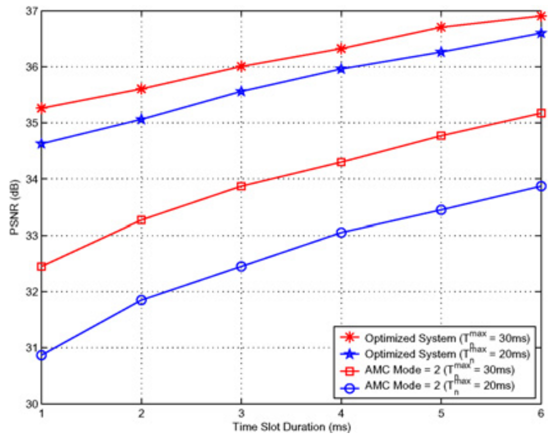


Fig. 8. Average PSNR comparison between the proposed quality-driven cross-layer optimized system and the existing system for real-time wireless video transmission over cognitive networks with various time slot durations T_0.

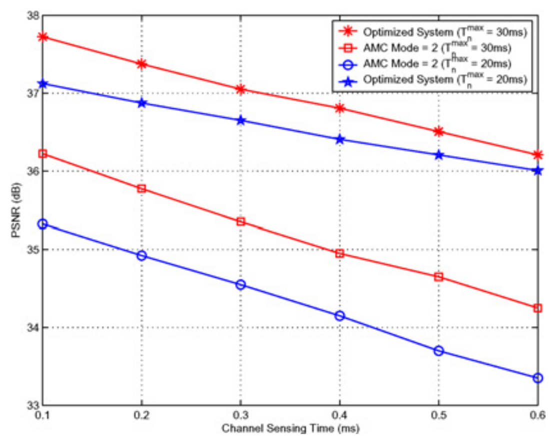


Fig. 9. Average PSNR comparison between the proposed quality-driven cross-layer optimized system and the existing system for real-time wireless video transmission over cognitive networks for various channel sensing times t_s.

respectively. From the figure, we can observe that the proposed system achieves higher performance gain when the time slot duration T_0 is smaller under the same packet delay bound T_n^{max}. Another observation is that when time slot durations T_0 are the same, higher performance gain can be achieved when channel sensing time t_s is set to 1 ms than that achieved by 0.5 ms. This indicates that under more stringent delay bound, the proposed system can adapt the time-varying channel and choose the optimal system parameters to achieve the best user-perceived video quality.

The effect of channel sensing time t_s on secondary users is also studied. Here, the time slot duration T_0 is set to 5 ms and SNR is 15 dB. The packet delay bound T_n^{max} is set to 20 ms and 30 ms, respectively. As shown in Fig. 9, by increasing t_s, the overall performance decreases. This is reasonable because when the time spent on channel sensing increases, the time spent on transmission ($T_0 - t_s$) decreases accordingly. However, the proposed system is able to achieve higher performance gain when more time is spent on channel sensing, thus minimizing the negative impact on the overall system performance.

In summary, all the experimental results demonstrate the significant performance enhancement of the proposed system for secondary users in CR networks. The experimental results also indicate that the performance gain is usually higher when the wireless channel experiences bad quality in a more stringent delay-bounded video application.

VII. Conclusion

In this paper, a cross-layer optimized system for real-time video transmission over CR networks has been proposed and studied. The proposed system can achieve the best possible video quality for secondary users, significantly improving the user experience of secondary users in CR networks, leading to the possibility of wide deployment of CR technologies in wireless video applications. The design problem is presented to minimize the expected video distortion under the constraint of packet delay bound, which has been formulated as a MIN-MAX problem and solved by dynamic programming. Experimental results have validated the effectiveness of the proposed system.

References

[1] D. Niyato and E. Hossain, "Cognitive radio for next-generation wireless networks: An approach to opportunistic channel selection in IEEE 802.11-based wireless mesh," *IEEE Wireless Commun.*, vol. 16, no. 1, pp. 46–54, Feb. 2009.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Selected Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] R. Tandra, A. Sahai, and S. M. Mishra, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, no. 5, pp. 824–848, Apr. 2009.

[4] H. Luo, D. Wu, S. Ci, A. Argyriou, and H. Wang, "Quality-driven TCP friendly rate control for real-time video streaming," in *Proc. IEEE GlobeCom*, Dec. 2008, pp. 1–5.

[5] H. Luo, A. Argyriou, D. Wu, and S. Ci, "Joint source coding and network-supported distributed error control for video streaming in wireless multi-hop networks," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1362–1372, Aug. 2009.

[6] Y. Yu, L. Wang, and Q. Yu, "Cross-layer architecture in cognitive ad hoc networks," *Commun. Mobile Comput.*, vol. 2, pp. 47–51, Jan. 2009.

[7] S. Ci and J. Sonnenberg, "A cognitive cross-layer architecture for next-generation tactical networks," in *Proc. IEEE MILCOM*, Oct. 2007, pp. 1–6.

[8] C. Ghosh and D. P. Agrawal, "ROPAS: Cross-layer cognitive architecture for wireless mobile ad hoc networks," in *Proc. Cog. Radio Orient. Wireless Netw. Commun.*, Aug. 2007, pp. 514–518.

[9] C. Ghosh, B. Xie, and D. P. Agrawal, "ROPAS: Cross-layer cognitive architecture for mobile UWB networks," in *Proc. Mobile Ad Hoc Sensor Syst.*, Oct. 2007, pp. 1–7.

[10] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Select. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[11] P. Mahonen, M. Petrova, J. Riihijarvi, and M. Wellens, "Cognitive wireless networks: Your network just became a teenager," in *Proc. IEEE INFOCOM*, Apr. 2006.

[12] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, document ITU-T Rec. H.264|ISO/IEC 14496-10 AVC, May 2003 [Online]. Available: ftp://ftp.imtc-files.org/jvt-experts/2003-03-Pattaya/JVT-G50r1.zip

[13] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[14] D. Wu, T. Hou, W. Zhu, H.-J. Lee, T. Chiang, Y.-Q. Zhang, and H. J. Chao, "On end-to-end architecture for transporting MPEG-4 video over the Internet," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 923–941, Sep. 2000.

[15] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error prone networks," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 952–965, Jun. 2000.

[16] A. Argyriou, "Real-time and rate-distortion optimized video streaming with TCP," *J. Image Commun.*, vol. 22, no. 4, pp. 374–388, Apr. 2007.

[17] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, Sep. 2004.

[18] H. Luo, D. Wu, and S. Ci, "Cross-layer optimized end-to-end TCP friendly rate control for real-time video streaming," *J. Vis. Commun. Image*, vol. 21, no. 2, pp. 98–106, Feb. 2010.

[19] H. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[20] Y. Sung, L. Tong, and H. V. Poor, "Neyman-Pearson detection of Gauss-Markov signals in noise: Closed-form error exponent and properties," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1354–1365, Apr. 2006.

[21] S. Akin and M. C. Gursoy, "Effective capacity analysis of cognitive radio channels for quality of service provisioning," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3354–3364, Nov. 2010.

[22] H. Wang, S. Tsaftaris, and A. K. Katsaggelos, "Joint source-channel coding for wireless object-based video communications utilizing data hiding," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2158–2169, Aug. 2008.

[23] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–673, Jul. 2003.

[24] M. S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer J. Wireless Commun.*, vol. 13, pp. 119–143, May 2000.

[25] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 172–180, May 2002.

[26] Z. Li, G. Schuster, and A. Katsaggelos, "MINMAX optimal video summarization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1245–1256, Oct. 2005.

[27] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*. Norwell, MA: Kluwer, 1997.

[28] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *Signal Process.*, vol. 15, pp. 23–50, Nov. 1998.

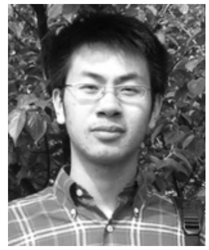
[29] Y. Andreopoulos, N. Mastronade, and M. van der Schaar, "Cross-layer optimized video streaming over wireless multi-hop mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2104–2115, Nov. 2006.

[30] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE J. Select. Areas Commun.*, vol. 25, no. 4, pp. 749–759, May 2007.

[31] E. Maani, P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1663–1671, Sep. 2008.

[32] H. Luo, S. Ci, D. Wu, and H. Tang, "Cross-layer design for real-time video transmission in cognitive wireless networks," in *Proc. IEEE INFOCOM Workshop Cognitive Wireless Commun. Netw.*, Mar. 2010, pp. 1–6.

[33] H. Luo, S. Ci, D. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE," *IEEE Commun.*, vol. 48, no. 2, pp. 102–109, Feb. 2010.



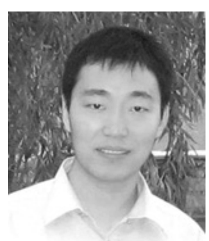
Haiyan Luo (S'09) received the B.S. degree from Dalian Jiaotong University (formerly Dalian Railway Institute), Dalian, China, and the M.S. degree from the Dalian University of Technology, Dalian. He is currently pursuing the Ph.D. degree from the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Lincoln.

His current research interests include multimedia networks and distributed systems.

Mr. Luo has served as a Referee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATION, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and several conferences, such as IEEE INFOCOM, IEEE GLOBECOM, and IEEE ICC.

Song Ci (S'98–M'02–SM'06) is an Associate Professor of computer and electronics engineering with the University of Nebraska-Lincoln, Lincoln. He is the Director of the Intelligent Ubiquitous Computing Laboratory, Omaha, NE, and holds a courtesy appointment of UNL Ph.D. in the Biomedical Engineering Program. He is also affiliated with the Chinese Academy of Sciences, Beijing, China, the Center for Advanced Surgical Technology at the University of Nebraska Medical Center, Omaha. His current research interests include dynamic complex system modeling and optimization, green computing and power management, dynamically reconfigurable embedded systems, content-aware quality-driven cross-layer optimized multimedia over wireless, cognitive network management and service-oriented architecture, and cyber-enabled e-healthcare.

He is an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and serves as a Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA Special Issue on "Quality-driven cross-layer design for multimedia communications." He is a Guest Editor of the IEEE NETWORK MAGAZINE Special Issue on "Wireless mesh networks: Applications, architectures and protocols" and the Associate Editor of four other international journals. He was the Co-Chair of the Technical Program Committee (TPC) of IEEE ICCCN in 2009 and the TPC member for numerous conferences. He is the Vice Chair of the Computer Society of the IEEE Nebraska Section. He is a member of the ACM and the ASHRAE. He is the Chair of the Special Interest Group, Quality-Driven Cross-Layer Design of the Multimedia Communications Technical Committee, IEEE Communications Society.



Dalei Wu (S'05) received the B.S. and M.E. degrees in electrical engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively, and the Ph.D. degree in computer engineering from the University of Nebraska-Lincoln, Lincoln, in 2010.

His current research interests include wireless multimedia networking, green computing, cyber-physical systems, and complex dynamic system modeling and optimization.