Cognitive Multihoming System for Energy and Cost Aware Video Transmission

Satyam Agarwal, *Student Member, IEEE* and Swades De, *Senior Member, IEEE*

***Abstract*—To alleviate the spectrum scarcity problem in the licensed cellular networks (LCN), we introduce a new paradigm called cognitive multihoming (CM), where a cognitive radio (CR)-enabled base station transmits to the users simultaneously over the licensed cellular bands and primary user bands in CR networks (CRNs). The CR aspect incurs lower cost, however at the expense of higher energy consumption due to intermittent channel sensing. On the other hand, transmission via LCN is expensive because of its licensing premium. To minimize the transmission cost while meeting the users energy and received video quality constraints, sensing duration and transmission rate over CRN, transmission rate over LCN, and network selection for retransmission of lost packets are adjusted. Solution to the multiuser resource allocation optimization problem is obtained by solving the cost minimization problem of a single-user system. The problem is nonconvex which is solved using convex-concave procedure. The proposed scheme is compared with the cases where a user operates over a single network, either LCN or CRN. The system performance results indicate that the proposed CM strategy significantly decreases the cost to the users as well as serves a higher number of users while maintaining the desired video quality and energy consumption constraints.**

***Index Terms*—Cognitive multihoming system, multi-radio clients, resource allocation, scalable video transmission, energy efficiency, convex-concave procedure**

I. Introduction

Explosive growth of Internet traffic is being experienced in recent years. Cisco research report in 2016 [1] indicated that the mobile video traffic demand is tipped to grow 11 times by 2020. It is anticipated that, soon the existing wireless networks may not be capable of delivering high-quality content to the mobile users despite the advances in wireless technology [2]. This is due to the licensed cellular networks (LCNs) lacking adequate spectrum availability. However, spectrum measurement studies have revealed that large parts of the licensed spectrum are underutilized [3]. Cognitive radio (CR) techniques can be employed to combat spectrum resource scarcity in the conventional cellular bands for mobile broadband quality-of-service (QoS) support.

On the receiver technology front, emergence of multihomed devices (multi-radio clients) [4] has enabled simultaneous access to multiple radio access technology in a heterogeneous network. Conventionally, multihoming capability of user devices is used for concurrent multipath transfer (CMT) [5] of data from remotely-located source via multiple networks. This approach invites challenges due to network dynamics, leading to selective packet loss over multiple paths and hence the user's difficulty in data decodability. In order to alleviate the spectrum scarcity problem, minimize the resource intensive video transmission load in LCNs, and quality degradation in conventional CMT via wireless networks, we propose a novel access paradigm, called cognitive multihoming (CM).

The authors are with the Department of Electrical Engineering and Bharti School of Telecom, Indian Institute of Technology Delhi, India (email: satyam6099@gmail.com; swadesd@ee.iitd.ac.in).

Preliminary version of this work was presented in IEEE GLOBECOM workshops 2015.

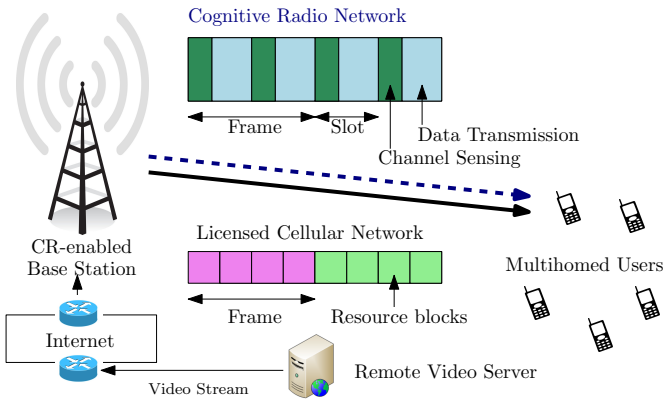


Fig. 1: Illustration of cognitive multihoming.

A. CM Concept Overview

As depicted in Fig. 1, in CM system the base station (BS) of LCN is also equipped with the CR functionality. The BS controller buffers the data packets from the video server across the Internet and divides the data stream suitably as per the available LCN and CR network (CRN) resources, and the BS transmits them to the user concurrently via LCN and CRN. The band over which the LCN and CRN operate are different. To meet dynamic demands of the real-time users, the service provider (SP) would want to keep its licensed band as idle as possible by making most use of CRN resource. To encourage usage of more CRN resources, data transmission over the CRN is offered at a discounted rate. This is in consonance with the fact that the cellular bands are costly, as the SP pays a huge premium for licensing. Operations over CRNs do not levy high cost to the users due to the opportunistic (secondary) usage of the primary user (PU) bands.

Transmission cost could be reduced if most of the content is transmitted via CRN. However, due to intermittent PU arrivals and imperfect channel sensing, there could be low throughput over CRN. Moreover, the channel needs to be sensed at regular intervals which further reduces the performance. Hence, reception over CRN can be characterized by low cost, low QoS, and low energy efficiency. On the other hand, transmissions

over LCN are marked by high cost and high QoS. Thus, there is an apparent cost-quality-energy tradeoff. The possible degradation in energy efficiency and QoS in CRN can be controlled in CM by optimally allocating resources from LCN and CRN as a function of a user's demand.

With the growing demand for mobile video streaming, there have been some works in the area of video streaming to multihomed devices [6], [7]. Different from these approaches, where the two separate end-to-end paths are utilized for video content delivery (as in CMT), in the present work we consider scalable video transmission to the users via CM technique. The users request unicast scalable video encoded content from the network. Due to high throughput demands of video services, transmissions via a single access network may not be sufficient for high quality content delivery. Instead, the users are equipped with multihomed devices so that high quality video content can be received. For battery operated devices, user satisfaction depends on the received video quality, device battery level, and the cost it pays for the content reception. To this end, in addition to transmission rate adaptation over single network [8], in the proposed CM approach, two novel techniques of video packet priority dependent sensing duration adaptation and network selection for packet retransmission are proposed to enhance user experience without causing PU performance degradation.

B. Key Objectives and Contributions

In the proposed CM architecture, we explore the optimally-shared scalable video transmission for enhanced user experience in terms of reception quality, energy savings, and cost.

The contributions of the paper are summarized as follows:

1) A novel CM framework is proposed, where the CR-enabled cellular BSs transmit optimally-divided user content over the two networks to the multihomed users (*Section III-A*).
2) Adaptation of transmission rate over LCN, sensing duration and transmission rate over CRN, and network selection for retransmission of lost packets is developed (*Section III-D*).
3) A joint optimization for maximizing the number of users served and minimizing the user cost is formulated, which accounts for the device energy and reception quality (*Section IV-A*).
4) An optimal algorithm for resource allocation and call admission control is proposed. The optimization problem being nonconvex, it is transformed into a difference of convex optimization problem and solved using the convex-concave procedure. The optimization problem is further decomposed to speed up the computation (*Section IV-B*).

Our system simulation results (*Section V*) demonstrate that, compared to single network access the CM system serves a higher number of users at a lower cost. In the next section, we discuss the related works.

II. Related Works

We survey the related works in three categories, namely, video transmission over single licensed wireless network, transmission over CRN, and transmission over multihomed networks.

A. Video Transmission over Single Wireless Network

Considerable work have been done on video transmission over single licensed wireless network. In [9], the authors proposed different retry limits to each layer of scalable video to improve quality of user experience (QoE). The works in [10], [11] proposed MDP based transmission scheduling of video packets over wireless fading channels. The authors in [11] further looked into energy-efficient scheduling of video packets. A link adaptation scheme for scalable video transmission over wireless local area network (WLAN) was proposed in [12]. The authors in [8] proposed energy minimization in video transmission by rate, power, and retransmission rate adaptation. Energy-efficient resource allocation for video transmission was proposed in [13]. Energy-efficient mobile video streaming was studied in [14]. The authors in [15] optimized high rate video transmission over wireless networks. We note that, while video transmission over single LCN has been considerably well investigated, scarcity of LCN resources to cater to ever-increasing volume of traffic and diverse user-end limitations and interests have motivated further research on radio resource utilization.

B. Video Transmission over CRN

There have been a few recent works reported on video transmission over CRN. In [16], the authors studied scalable video multicast over CRN. The work in [17] studied the impact of spectrum sensing frequency on video transmission. In [18], sensing delays in real-time video transmission was addressed by breaking the sensing duration into small time slots. In [19], the authors considered scalable video coding and transmission rate adaptation jointly, for energy-efficient video streaming over CR. A scheme to reduce unnecessary sensing and extend the transmission times for optimal SU operation over multiple PU channels was presented in [20]. QoS based resource allocation for CRN was proposed in [21]. In [22], the authors proposed a QoE-driven channel allocation scheme for CR users. A multiuser video streaming scheme in cellular CRNs was presented in [23], where spectrum sensing, power allocation, and channel assignment strategies were jointly optimized to maximize QoS. The difficulties in providing service to CR users in delay sensitive applications due to intermittent PU activity has motivated us to consider access strategy via multiple networks simultaneously.

C. Video Transmission to Multihomed Clients

In heterogeneous networks, the users are capable of accessing multiple RANs. They can access multiple RANs simultaneously as in multihoming, or they may select a single network for their operation. Simultaneous video transmission over multiple networks has drawn considerable research attention in recent years. Different from our proposed model where the data stream is split at the last hop, the works in the literature on multihoming consider different paths from

content server to the user device. Hence, their main focus remain on transport layer congestion control over multiple networks. The studies in [4], [5] considered CMT based video delivery and proposed a TCP based solution for efficient video reception. Systematic Raptor codes were adopted in [6] to mitigate video quality degradation caused by wireless channel in multipath video transmission. The authors in [24] analyzed probabilistic multipath transmission of streaming video packets over multiple networks via Markov chains. High definition video streaming over heterogeneous networks was considered in [7], [25]. These studies did not account cost and energy consumption related issues.

In [26], energy-efficient video transmission over multi-homed network was studied. An MDP based cost minimization framework for video transmission was presented in [27]. The work in [28] reported performance optimization in terms of cost, signaling load, and session quality, for media streaming over heterogeneous networks. The inter-cell network cooperation in [29] implicitly used the multihoming concept in a homogeneous LCN scenario. A survey on bandwidth aggregation techniques in heterogeneous multi-homed devices was presented in [30]. All these studies did not consider energy consumption and user cost jointly. The approach in [31] is closest to our proposed model, where the authors studied user cost and energy optimized video transmission over two access networks. However, it did not consider CR-enabled multihoming aspects. Additionally, layer based parameter adaptation in scalable video streaming and admission control in multiuser scenario was not in its scope of study.

While LTE-U [32] has been introduced to enhance the capacity of LTE by extending its operation over unlicensed bands, inter-operability with other devices in the unlicensed bands pose a big challenge. CM offers an alternate to LTE-U wherein the opportunistic use of licensed bands along with the LCN enhances the system capacity. Moreover, transmission power limitations over the unlicensed bands restricts LTE-U operation regions. In contrast, CR-enabled CM easily adapts itself to different environments by employing spectrum sensing mechanisms which is difficult in LTE-U operations.

As an advance to the existing body of works, the current work in this paper distinguishes itself on the following aspects: *(i) The presented CM system architecture is a novel approach to efficient and cost-effective utilization of wireless spectrum resource toward QoS provisioning, and it is different from the* conventional *multihoming and CMT. (ii) The usage of CRN resources opportunistically in multihoming approach, which can effectively integrate pure Licensed (white-space) Spectrum Access and Dynamic Spectrum Access, has not been addressed before. (iii) Consideration of cost to the users and device energy consumption for scalable video streaming over the heterogeneous networks has not been jointly dealt with in the literature.* The basic CM model was presented recently in [33]. It however did not consider the application-specific details, the related cost, and energy optimization schemes.

III. System Model and Preliminaries

We now introduce the CM concepts along with the system assumptions, video traffic model, and user device characteristics. Single network access (SNA) is also briefly outlined for comparison. Major notations used in this paper are listed in Table I.

TABLE I: List of major notations and their descriptions.

B_{rb} and t_{rb}	Bandwidth and duration of a resource block (RB)
B_{cr} and t_{cr}	Bandwidth and duration of a slot over CRN
p_{cr}	PU channel idling probability in a slot
$p_f(\delta)$	Probability of false alarm with sensing duration δ
$p_m(\delta)$	Probability of misdetection with sensing duration δ
$\mathcal{F}_{lc}(z)$, $\mathcal{F}_{cr}(r)$	Probability of packet failure transmitted at rate z/r over LCN/CRN due to fading and path loss
z_l, r_l	Transmission rate over LCN/CRN for lth layer
δ_l	Sensing duration for lth layer transmission
$\mathcal{T}$	Group of pictures (GOP) duration
$J(\delta)$	Transmission probability over a slot in CRN with sensing duration δ
Q_{th}	Minimum received video quality threshold
E_{th}	Maximum energy consumption threshold
α	Per packet overhead (bits)
$\mathcal{N}_{rb}$	Number of RBs available per GOP
$p_{lc,l}^{rtx}$	Retransmission probability of lth layer content over LCN that is originally transmitted over LCN
$p_{cr,l}^{rtx}$	Retransmission probability of lth layer content over CRN that is originally transmitted over CRN
n_l	Maximum number of retransmissions for lth layer content
d_l	lth layer video data content (bits) in a GOP duration
u_l	lth layer content (bits) transmitted over CRN in a GOP
$w_{lc,l}$, $w_{cr,l}$	Amount of lth layer content (including retransmissions) transmitted over LCN/CRN in a GOP duration
$\mathcal{G}_{cr}(\delta, r)$	Probability of packet failure transmitted at rate r and with sensing duration δ over a slot in CRN
ϕ_{lc}, ϕ_{cr}	Cost per unit time of LCN/CRN usage
$\xi_{lc,rx}$, $\xi_{cr,rx}$	Reception power consumption over LCN/CRN
$\xi_{lc,ta}$, $\xi_{cr,ta}$	Tail power consumption over LCN/CRN
p_m^{th}	Maximum allowable probability of misdetection

A. Cognitive Multihoming

In CM (Fig. 1), apart from operating over the licensed bands, the cellular BSs have an in-built functionality to operate over the CR bands. The network operates on at least one LCN and one CRN band. LCN could be from, e.g., LTE and WCDMA standards, while CRN could be based on the IEEE standards, such as, 802.11af and 802.22. Downlink video content transmission is considered. The video content is buffered at the BS from the external server. The BS decides on how the data should be split across the two networks for optimal operation. As a natural advantage of multihoming, CM avoids any communication overhead for data splitting between the user and external server. However, unlike in CMT, out of order delivery of packets across the two networks is minimized by optimal data stream splitting at the BS. Thus, the drawback of congestion related delay variations in conventional multihoming is not present in CM.

LCN and CRN system specific details are presented next.

1) Transmission over LCN: The resources over LCN are divided into time-frequency resource blocks (RBs). The bandwidth and time duration of an RB are respectively denoted as B_{rb} and t_{rb}. Over an RB, transmitter rate depends on the channel conditions. Block fading is considered where the channel gain across an RB remains constant over the duration t_{rb}. Denoting $P_{lc,tx}$ as the BS transmit power, received signal

power $P_{lc,rx}$ at a user located $\mathcal{L}$ distance away from the BS is given by the Friis formula:

$$P_{lc,rx} = P_{lc,tx} G_T G_R \left(\frac{\lambda}{4\pi\mathcal{L}}\right)^{\eta}$$

where G_T and G_R are respectively the transmitter and receiver antenna gains, λ is the transmitted signal wavelength, and η is the path loss coefficient. Considering Rayleigh fading (with 0 mean and unit variance) between the BS and user, the rate offered over an RB is given by Shannon's capacity formula $Z = B_{rb}\log_2\left(1+\frac{h_{lc}P_{lc,rx}}{\sigma^2_{lc,n}}\right)$, where h_{lc} is the channel fading power gain and $\sigma^2_{lc,n}$ is the noise variance. Outage probability at transmission rate z bps is given as:

$$\mathcal{F}_{lc}(z) = Pr(Z \leq z) = 1 - exp\left(-\frac{\sigma^2_{lc,n}(2^{z/B_{rb}}-1)}{P_{lc,rx}}\right). \tag{1}$$

2) Transmission over CRN: CRNs operate opportunistically over the licensed PU channels. We denote, bandwidth B_{cr} of a PU channel can be used by a CR-user at a time. Time is divided into slots of duration t_{cr} units. PUs transmission over the channel is considered to follow ON-OFF model [34]. The probability of PU channel being idle in a slot is denoted as p_{cr}. Over a slot, the CR-enabled BS senses the channel for a duration δ, and if found idle, it transmits over the channel for the remaining time $t_{cr} - \delta$. Relative to the PU coverage area, the users and the BS are considered co-located in a small geographical area, such that they experience similar channel conditions. Probabilities of false alarm $p_f(\delta)$ and misdetection $p_m(\delta)$ as a function of sensing duration δ are respectively given as [35]:

$$p_f(\delta) = \text{erfc}\left(\left(\frac{\epsilon}{\sigma^2_{cr,u}} - 1\right)\sqrt{\delta f_s}\right) \tag{2}$$

$$p_m(\delta) = \text{erfc}\left(\left(\frac{\epsilon}{\sigma^2_{cr,u}} - \gamma - 1\right)\sqrt{\frac{\delta f_s}{2\gamma+1}}\right) \tag{3}$$

where $\sigma^2_{cr,u}$ is the noise power variance, f_s is the channel sampling frequency, ϵ is the sensing threshold, and γ is the PU signal SNR at the CR node. The user data is transmitted at a rate r bps over the remaining part of the slot if the channel is sensed idle. Packet transmission probability over CRN in a slot is given as:

$$J(\delta) = (1-p_{cr})p_m(\delta) + p_{cr}(1-p_f(\delta)). \tag{4}$$

Similar to (1), outage probability over the PU channel in CRN due to channel fading is:

$$\mathcal{F}_{cr}(r) = 1 - exp\left(-\frac{\sigma^2_{cr,n}(2^{r/B_{cr}}-1)}{P_{cr,rx}}\right) \tag{5}$$

where $\sigma^2_{cr,n}$ is the channel noise variance and $P_{cr,rx}$ is the received power over the CRN. Considering energy detection based spectrum sensing, for accounting interference to a CR transmission, all other transmissions are combined together.

These developments on transmission error over LCN and CRN will be used in Section IV.

B. Video Traffic Model

We now outline the traffic characteristics and video quality metric. We consider scalable video coding (SVC) video encoder that allows graceful degradation of video quality caused by wireless channel fading. It consists of a base layer (layer 1) and several enhancement layers (layers $2, 3, \cdots, L$). Layer l_1 has a higher priority over layer l_2 if $l_1 < l_2$. Base layer provides the basic quality level of the video and it is decoded independently of the higher layers. Enhancement layer l_1 can only be decoded once all the layers $l < l_1$ are decoded. Each layer is independently encoded at a specific rate.

A predictive model of QoE for Internet video was presented in [36]. However, considering the SVC video transmission in our work, we define received video quality metric similarly as in [4], [12]. Distortion of H.264/SVC encoded video can be quantified as $\mathcal{D} = \mathcal{D}_{enc} + \mathcal{D}_{loss}$. $\mathcal{D}_{enc}$ is the average encoding distortion, given as $\mathcal{D}_{enc} = D_0 + \frac{\theta_0}{X - R_0}$, where X is the average bit rate, and the parameters D_0, θ_0, and R_0 are constants that depend on the encoder and video characteristics. Packet loss induced distortion $\mathcal{D}_{loss}$ is independent of $\mathcal{D}_{enc}$. Denoting φ_l as the probability of lth layer content loss, $\mathcal{D}_{loss}$ can be expressed as:

$$\mathcal{D}_{loss} = \sum_{l=1}^{L} \varphi_l \Omega_l, \quad \text{where } \Omega_l = \left(\frac{\theta_0}{X_{l-1} - R_0} - \frac{\theta_0}{X_l - R_0}\right). \tag{6}$$

Ω_l is the weight associated with the lth layer, and X_l is the rate of the l layers video content.

Video quality metric Q is defined as the distortion impact of the packets received correctly to the total available packets [26]. It is given as:

$$Q = \frac{\sum_{l=1}^{L}\Omega_l(1-\varphi_l)}{\sum_{l=1}^{L}\Omega_l} \cdot 100\%. \tag{7}$$

Video stream is divided into group of pictures (GOP) with each GOP duration of $\mathcal{T}$ units.

C. User Device Operation

Before presenting adaptive video transmission, we briefly discuss how the users (multihomed clients) operate. A user device is enabled with multi-RAT (radio access technology) access functionality by which it can simultaneously operate over the LCN and CRN. All user devices are considered battery operated and therefore their respective reception processes are sensitive to the remaining battery energy. Part of this energy consumption is due to reception of the video content over the two networks. Consequently, the upper bound on energy consumption and the quality of video (number of layers) requested is determined by its remaining battery level.

As noted in the experimental studies in [37], [38], energy consumption in mobile devices due to network interface is mainly on three fronts, namely, ramp, tail, and data transfer energy. Ramp energy is consumed when the radio transits from *idle* state to *transmit* state. Data transfer energy is proportional to the duration of time the radio is transmitting/receiving. Tail energy refers to the lingering of the radio at high energy

state during the inactivity period after the transmission is over, which is on the order of tens of seconds [37]. The network transmits to the device for a fraction of time in a GOP. Thus, reception energy is consumed for the time device is actually receiving, and the tail energy is consumed when the device is idle in a GOP. The device does not incur ramp energy consumption, as the radio mostly remains in the high energy state during the reception.

Along with meeting the energy consumption constraints, the network should ensure high quality video reception to the user end at a low cost. The communication cost (cost charged to the user) is proportional to the amount of resources allocated via the two networks. A user is more satisfied when it receives the video content at a lower cost while satisfying its energy consumption constraint.

In the following, we discuss adaptive video transmission to a CM-capable user.

D. Video Transmission to CM User

The BS transmits video stream to a user optimally across the two networks such that the user's cost is minimized subject to the user's energy consumption and video quality constraints. Note that, the user cost minimization by SP does not affect its revenue maximization objective, as the profit to SP may be completely different over the two networks. Video content over a GOP and SVC video parameters are considered available at the BS before the GOP starts. Consider that the BS is capable of performing layer based video transmission. It can form packets of desired length and transmits them to a user. A lost packet is retransmitted by the BS (based on acknowledgement (ACK) notification via independent LCN RBs) to the user up to a layer-dependent predefined number (n_l) of retransmission attempts. The undelivered packets within a GOP duration are dropped, resulting in distortion. For enhanced user experience, transmission rate adaptation over the LCN, sensing duration and transmission rate adaptation over the CRN, and number of retransmission attempts and network selection for retransmission are optimized for each layer of video transmission. We describe these three mechanisms next.

1) Transmission rate adaptation over LCN/CRN: According to (1) and (5), a higher transmission rate has a lower probability of success. However, higher transmission rate introduces lower cost as well as energy consumption to the user. Hence, while the lower layers (high distortion impact) are transmitted at a lower rate so that their correct reception probability is higher, the higher layers are transmitted at a higher rate.

2) Sensing duration adaption over CRN: Sensing duration δ over a slot plays a key role in determining the reception performance. With a small δ, probabilities of misdetection and false alarm are large, but the available transmission duration are also longer. Therefore, for balanced cost and quality, for the different priority of video packets δ can be accordingly chosen.

3) Cross network retransmissions: Video packets transmitted can be lost due to path loss, channel fading, or channel sensing misdetections. Such packets are retransmitted within the allowable GOP time window to reduce distortion at the receiver. Network selection for retransmission is also critical. Retransmission over CRN could reduce cost, however at a higher energy consumption. If the retransmission request for the lth layer content initially transmitted over LCN is received by the BS, it can either retransmit the content over LCN or CRN.

E. Benchmark Schemes

For comparison purpose, we consider two schemes. In the first, we consider that the user is associated with only a single network, either LCN or CRN. We call it as single network access (SNA)-LCN or SNA-CRN. If the user is associated with LCN, transmission rate is optimized for its optimal performance. In SNA-CRN, sensing duration and transmission rate adaptation are performed to guarantee optimal performance. To explicitly demonstrate the benefit of layer-dependent parameter adaptation in the proposed CM (we call it as optimized CM, or CM-opt), the comparative scheme we consider is the basic CM (CM-basic). In CM-basic, transmission rate adaptation over the LCN and CRN, sensing time adaptation, and number of retransmission and network selection for retransmission are optimally chosen for all video packets without distinguishing their respective priority levels, to obtain low cost transmission to the users.

IV. Problem Formulation and Solution Procedure

We analyze the multiuser scenario, where a fixed pool of total channel resource is traded with multiple user demands of video content delivery. First, the multiuser resource allocation optimization problem is formulated. Subsequently, to achieve solution of the multiuser optimization, for an individual user, single-user cost minimization problem is solved. Delay-tolerant traffic requests are not considered in our current study, as it does not require resource guarantee.

A. Problem Formulation

Let $\mathbb{N}$ be the number of users requesting video content. Resources are allocated to the users per GOP. Denote $\mathcal{N}_{rb}$ as the maximum number of RBs available to the users in a GOP. Due to limited resource availability at the BS, some of the user requests may not be fulfilled, leading to their non-admission to the network. Resource allocation by the BS is such that the user cost is minimized while the number of users served is maximized.

For a user requesting L layers of video content, let d_l amount of data (in bits) be transmitted for the lth layer content over a GOP. The BS transmits data of layer l at a rate z_l over LCN and at a rate r_l over CRN. We denote the sensing duration for layer l packet as δ_l. Also, denote that the lth layer content can be retransmitted at most $\lceil n_l \rceil$ times, where n_l can take a fractional value in general. It is upper bounded by n_{max}, which is the maximum number of retransmissions allowed for any individual packet within a GOP. Let $p_{lc,l}^{rtx}$ (respectively, $p_{cr,l}^{rtx}$) be the probability that the retransmission of lth layer content transmitted initially over LCN (respectively, CRN) is carried out over LCN (respectively, CRN) itself.

As the data stream d_l is split across the two networks, let u_l amount be initially transmitted over CRN. The remaining data volume $(d_l - u_l)$ is initially transmitted over LCN. To compute the total traffic over LCN for the lth layer content we note that, due to wireless channel uncertainties, part of the failed data is retransmitted over LCN (with probability $p_{lc,l}^{rtx}$). The remaining volume of retransmission is over CRN, with probability $(1 - p_{lc,l}^{rtx})$. Given the maximum number of retransmissions n_l for the lth layer, $\sum_{i=0}^{n_l} \mathcal{F}_{lc}(z_l)^i = \frac{1-\mathcal{F}_{lc}(z_l)^{n_l+1}}{1-\mathcal{F}_{lc}(z_l)}$ is the expected number of transmission attempts carried out including the initial transmission for the part of data retransmitted over LCN. Hence, for transmitting $(d_l - u_l)$ data over LCN, the expected data transmission over LCN is $(d_l - u_l)\left(p_{lc,l}^{rtx}\left(\frac{1-\mathcal{F}_{lc}(z_l)^{n_l+1}}{1-\mathcal{F}_{lc}(z_l)}\right) + 1 - p_{lc,l}^{rtx}\right)$. The data transmitted over CRN could also be lost due to misdetection in channel sensing, path loss, and channel fading. The probability of packet transmission failure $\mathcal{G}_{cr}(\delta_l, r_l)$ over CRN is given as:

$$\mathcal{G}_{cr}(\delta_l, r_l) = (1 - p_{cr})p_m(\delta_l) + p_{cr}(1 - p_f(\delta_l))\mathcal{F}_{cr}(r_l). \quad (8)$$

Part of the lost data is retransmitted over LCN. Out of the original share of transmitted data u_l over CRN, the amount of data retransmitted over LCN is $u_l(1 - p_{cr,l}^{rtx})\mathcal{G}_{cr}(\delta_l, r_l)$. The expected number of retransmissions over LCN is $\sum_{i=0}^{n_l-1} \mathcal{F}_{lc}(z_l)^i = \frac{1-\mathcal{F}_{lc}(z_l)^{n_l}}{1-\mathcal{F}_{lc}(z_l)}$ for the data retransmitted over LCN. Overall, the total amount of data transmitted over LCN for the lth layer in a GOP is:

$$\begin{aligned} w_{lc,l} =& (d_l - u_l)\left(p_{lc,l}^{rtx}\left(\frac{1 - \mathcal{F}_{lc}(z_l)^{n_l+1}}{1 - \mathcal{F}_{lc}(z_l)}\right) + 1 - p_{lc,l}^{rtx}\right) \\ &+ u_l(1 - p_{cr,l}^{rtx})\mathcal{G}_{cr}(\delta_l, r_l)\frac{1 - \mathcal{F}_{lc}(z_l)^{n_l}}{1 - \mathcal{F}_{lc}(z_l)}. \end{aligned} \quad (9)$$

Similarly, total lth layer data transmission over CRN in a GOP is obtained as:

$$\begin{aligned} w_{cr,l} =& u_l\left(p_{cr,l}^{rtx}\left(\frac{1 - \mathcal{G}_{cr}(\delta_l, r_l)^{n_l+1}}{1 - \mathcal{G}_{cr}(\delta_l, r_l)}\right) + 1 - p_{cr,l}^{rtx}\right) \\ &+ (d_l - u_l)(1 - p_{lc,l}^{rtx})\mathcal{F}_{lc}(z_l)\frac{1 - \mathcal{G}_{cr}(\delta_l, r_l)^{n_l}}{1 - \mathcal{G}_{cr}(\delta_l, r_l)}. \end{aligned} \quad (10)$$

Probability of packet transmission failure for the data transmitted (including retransmission) via LCN is $\mathcal{F}_{lc}(z_l)^{n_l+1}$, while that for the data transmitted over LCN and retransmitted over CRN is $\mathcal{F}_{lc}(z_l)\mathcal{G}_{cr}(\delta_l, r_l)^{n_l}$. Hence, for $(1 - u_l/d_l)$ fraction of lth layer content initially transmitted over LCN, probability of failure after retransmissions is $(p_{lc,l}^{rtx}\mathcal{F}_{lc}(z_l)^{n_l+1} + (1 - p_{lc,l}^{rtx})\mathcal{F}_{lc}(z_l)\mathcal{G}_{lc}(\delta_l, r_l)^{n_l})$. Similarly for u_l/d_l fraction of the lth layer content, the probability of transmission failure is $\left(p_{cr,l}^{rtx}\mathcal{G}_{lc}(\delta_l, r_l)^{n_l+1} + (1 - p_{cr,l}^{rtx})\mathcal{G}_{lc}(\delta_l, r_l)\mathcal{F}_{lc}(z_l)^{n_l}\right)$. Consolidating, probability of lth layer content loss is given as:

$$\begin{aligned} \varphi_l =& (1 - \frac{u_l}{d_l})\left(p_{lc,l}^{rtx}\mathcal{F}_{lc}(z_l)^{n_l+1} + (1 - p_{lc,l}^{rtx})\mathcal{F}_{lc}(z_l)\mathcal{G}_{lc}(\delta_l, r_l)^{n_l}\right) \\ &+ \frac{u_l}{d_l}\left(p_{cr,l}^{rtx}\mathcal{G}_{lc}(\delta_l, r_l)^{n_l+1} + (1 - p_{cr,l}^{rtx})\mathcal{G}_{lc}(\delta_l, r_l)\mathcal{F}_{lc}(z_l)^{n_l}\right). \end{aligned} \quad (11)$$

The overall received video quality is quantified using (7).

A packet is transmitted per RB/slot with an overhead of α bits. From (9), the amount of LCN RBs required by the user is given as:

$$N_{rb} = \sum_{l=1}^{L} \frac{w_{lc,l}}{(z_l t_{rb} - \alpha)} \quad (12)$$

while the spectrum leasing duration over CRN using (10) is:

$$T = t_{cr} \sum_{l=1}^{L} \frac{w_{cr,l}}{J(\delta_l)(r_l(t_{cr} - \delta_l) - \alpha)}. \quad (13)$$

Over LCN, cost of communication charged to the user is proportional to the number of RBs allocated to the user. Denoting cost charged per unit time as ϕ_{lc}, cost to the user over LCN is:

$$\mathcal{C}_{lc} = \phi_{lc} t_{rb} \sum_{l=1}^{L} \frac{w_{lc,l}}{(z_l t_{rb} - \alpha)}. \quad (14)$$

Over the CRN, cost to the user is considered proportional to the time duration the CR spectrum is used for its service. Let ϕ_{cr} be the cost to the user per unit time. Thus, the user cost over CRN is:

$$\mathcal{C}_{cr} = \phi_{cr} t_{cr} \sum_{l=1}^{L} \frac{w_{cr,l}}{J(\delta_l)(r_l(t_{cr} - \delta_l) - \alpha)}. \quad (15)$$

Total cost to the user is $\mathcal{C} = \mathcal{C}_{lc} + \mathcal{C}_{cr}$.

Energy consumption on account of reception over LCN is:

$$\Xi_{lc} = (\xi_{lc,rx} - \xi_{lc,ta}) t_{rb} \sum_{l=1}^{L} \frac{w_{lc,l}}{(z_l t_{rb} - \alpha)} + \xi_{lc,ta}\mathcal{T} \quad (16)$$

where reception and tail power consumption over LCN are respectively $\xi_{lc,rx}$ and $\xi_{lc,ta}$. Over CRN, the device remains in receive mode in the entire slot duration. Denoting receive and tail power consumption over CRN respectively as $\xi_{cr,rx}$ and $\xi_{cr,ta}$, total energy consumption in reception over CRN is:

$$\Xi_{cr} = (\xi_{cr,rx} - \xi_{cr,ta}) t_{cr} \sum_{l=1}^{L} \frac{w_{cr,l}}{J(\delta_l)(r_{cr}(t_{cr} - \delta_l) - \alpha)} + \xi_{cr,ta}\mathcal{T}. \quad (17)$$

For the multiuser resource allocation optimization formulation, we use superscript k over a variable to denote it for the kth user. Depending on the remaining battery level, the kth user requests $L^{(k)}$ layers of the video content at a minimum quality threshold $Q_{th}^{(k)}$ and a maximum energy consumption constraint $E_{th}^{(k)}$. Let $\psi^{(k)} = 1$ (respectively, 0) denote the user k served (respectively, not served). A reward β is associated with each user admitted. The resource allocation optimization problem is formulated in (18).

The optimization problem (18) jointly maximizes the number of users served and minimizes the cost of video transmission to the users while ensuring their demanded quality (cf. C1) and energy consumption (cf. C2) constraints. To protect the PUs from interference due to CR activity, the sensing duration is chosen such that the probability of misdetection is upper-bounded by p_m^{th} (cf. C3). C4 bounds the number of packet retransmissions to n_{max}, while C5 upper bounds u_l to

$$(P1) \quad \mathcal{C}^*_{net} = \text{ minimize} \sum_{k=1}^{\mathbb{N}} (\mathcal{C}^{(k)} - \beta)\psi^{(k)} \tag{18}$$

$$\text{s.t.} \quad C1: \frac{\sum_{l=1}^{L^{(k)}} \Omega_l^{(k)}(1-\varphi_l^{(k)})}{\sum_{l=1}^{L^{(k)}} \Omega_l^{(k)}} \cdot 100\% \geq Q_{th}^{(k)}, \quad \forall\ k \in 1,2,\cdots,\mathbb{N},$$

$$C2: \Xi_{lc}^{(k)} + \Xi_{cr}^{(k)} \leq E_{th}^{(k)}, \forall\ k \in 1,2,\cdots,\mathbb{N},$$

$$C3: p_m^{(k)}(\delta_l) \leq p_m^{th},\ \forall\ l \in 1,2,\cdots,L^{(k)}, \quad \forall\ k \in 1,2,\cdots,\mathbb{N},$$

$$C4: n_l^{(k)} \leq n_{max},\ \forall\ l \in 1,2,\cdots,L^{(k)}, \quad \forall\ k \in 1,2,\cdots,\mathbb{N},$$

$$C5: u_l^{(k)} \leq d_l^{(k)},\ \forall\ l \in 1,2,\cdots,L^{(k)}, \quad \forall\ k \in 1,2,\cdots,\mathbb{N},$$

$$C6: \sum_{k=1}^{\mathbb{N}} N_{rb}^{(k)}\psi^{(k)} \leq \mathcal{N}_{rb},$$

$$C7: \sum_{k=1}^{\mathbb{N}} T^{(k)}\psi^{(k)} \leq \mathcal{T},$$

$$C8: \psi^{(k)} \in \{0,1\},\ \forall\ k \in 1,2,\cdots,\mathbb{N}.$$

the lth layer content d_l. C6 and C7 are the resource availability constraints respectively over LCN and CRN.

Due to the integer variable $\psi^{(k)}$, (P1) in (18) is a mixed-integer optimization problem. For individual users, the factor $J(\delta_l)$ appearing in energy (cf. (17)), cost (cf. (15)), and CRN resource (cf. (13)) constraints is nonconvex in δ_l. The terms $\mathcal{F}_{lc}(z_l)$ and $\mathcal{G}_{cr}(\delta_l, r_l)$ used in computation of $w_{cr,l}$ (cf. (9)) and $w_{lc,l}$ (cf. (10)) are also nonconvex. Hence, (P1) is a mixed-integer nonconvex optimization problem. The solution to this problem is proposed next.

B. Solution to the Optimization Problem

A higher number of users can be served if the users are allocated resources in the increasing order of their demands. We note from (14) and (15) that the cost to a user is directly proportional to the resource allocated. For each user, we independently compute the optimal resource requirement and corresponding optimal cost to the user (cf. Section IV-B1). A user with lower cost budget is served first, subject to resource availability. As noted later in Section V-A1, the users with stronger energy constraint are allocated resources via LCN, as they cannot be supported by CRNs due to high energy consumption in CRNs. Thus, if for a user the available LCN RBs are less than the required RBs, the user would not be admitted in to the network. On the other hand, if the available CRN resource are less than the required CRN resource for a user, then optimal resource requirement for the user is recomputed after updating the CRN resource constraint. The remaining users are again sorted in the increasing order of their optimal cost and the allocation process continues. The proposed algorithm is presented in Algorithm 1, which provides an optimal solution as it simultaneously achieves low cost to the users while maximizing the number of users served.

Algorithm 1: Resource allocation in multiuser scenario.

1. Obtain optimal resource requirements for each user k. Optimal cost to the user, number of LCN RBs, and PU channel leasing duration are $\mathcal{C}^{(k)*}$, $N_{rb}^{(k)*}$, and $T^{(k)*}$, respectively;
2. Arrange the users in the increasing order of their optimal cost $\mathcal{C}^{(k)*}$;
3. Set the available resources: $N_{rb}^a \leftarrow \mathcal{N}_{rb}$ and $T^a \leftarrow \mathcal{T}$ and $j = 1$;
while $j \leq \mathbb{N}$ **do**
  if $N_{rb}^{(j)*} \leq N_{rb}^a$ *and* $T^{(j)*} \leq T^a$ **then**
    4. Allocate resources to the jth user;
    5. Update the available resources $N_{rb}^a \leftarrow N_{rb}^a - N_{rb}^{(j)*}$ and $T^a \leftarrow T^a - T^{(j)*}$;
    6. $j \leftarrow j + 1$;
  else
    if $N_{rb}^{(j)*} > N_{rb}^a$ **then**
      7. Drop the user ;
      8. $j \leftarrow j + 1$;
    else
      9. Recompute optimal resource requirements with available CRN resource constraint $T^{(j)} \leq T^a$;
      10. Sort the remaining users (including the jth user) again in the increasing order of their optimal cost $C^{(j)*}$;
    end
  end
end

1) Resource allocation optimization to single user: In steps 1 and 9 of Algorithm 1, the optimal resource and cost to the individual users are required. To obtain this, we need to solve a cost minimization problem for a single user. The corresponding optimization problem (P2) is formulated in (19), where superscript k (to indicate the kth user) is omitted for brevity.

N_{rb}^a in C6a and T^a in C7a are the available resources over LCN and CRN, respectively. C9 and C10 are the constraints corresponding to the amount of lth layer content transmitted over LCN ($w_{lc,l}$) and CRN ($w_{cr,l}$), respectively.

As stated in Section IV-A, the expressions for cost, energy, and video quality metrics in (P1) are nonconvex. Hence, (P2) is a nonconvex optimization problem. In the following, the problem is transformed to a difference of convex (DC) optimization problem and solved using the convex-concave procedure (CCP). Before proceeding further, we briefly discuss the DC optimization problem and introduce some useful lemmas.

2) DC optimization problem: A DC optimization problem can be written in the following form:

$$(P3) \ \underset{\mathbf{x}}{\text{minimize}} \quad f_0(\mathbf{x}) - g_0(\mathbf{x}) \qquad \text{s.t.} \quad f_i(\mathbf{x}) - g_i(\mathbf{x}) \leq 0, \forall\ i \in 1,2,\cdots M \tag{20}$$

$$(P2) \quad \mathcal{C}^* \triangleq \mathcal{C}^*_{lc} + \mathcal{C}^*_{cr} = \underset{\substack{z_l,\delta_l,r_l,u_l,n_l \\ p^{rtx}_{lc,l},p^{rtx}_{cr,l}}}{\text{minimize}} \sum_{l=1}^{L} \left(\phi_{lc} t_{rb} \frac{w_{lc,l}}{(z_l t_{rb} - \alpha)} + \phi_{cr} t_{cr} \frac{w_{cr,l}}{J(\delta_l)(r_l(t_{cr} - \delta_l) - \alpha)} \right) \tag{19}$$

s.t. $C1{-}C5$ from (18) (superscript k removed for brevity),

$$C6a : \sum_{l=1}^{L} \frac{w_{lc,l}}{(z_l t_{rb} - \alpha)} \leq N^a_{rb},$$

$$C7a : \sum_{l=1}^{L} \frac{w_{cr,l}}{J(\delta_l)(r_l(t_{cr} - \delta_l) - \alpha)} \leq \frac{T^a}{t_{cr}},$$

$$C9 : (d_l - u_l) \left(p^{rtx}_{lc,l} \left(\frac{1 - \mathcal{F}_{lc}(z_l)^{n_l+1}}{1 - \mathcal{F}_{lc}(z_l)} \right) + 1 - p^{rtx}_{lc,l} \right) + u_l (1 - p^{rtx}_{cr,l}) \mathcal{G}_{cr}(\delta_l, r_l) \frac{1 - \mathcal{F}_{lc}(z_l)^{n_l}}{1 - \mathcal{F}_{lc}(z_l)} = w_{lc,l}, \quad \forall\ l \in 1, 2, \cdots, L,$$

$$C10 : u_l \left(p^{rtx}_{cr,l} \left(\frac{1 - \mathcal{G}_{cr}(\delta_l, r_l)^{n_l+1}}{1 - \mathcal{G}_{cr}(\delta_l, r_l)} \right) + 1 - p^{rtx}_{cr,l} \right) + (d_l - u_l)(1 - p^{rtx}_{lc,l}) \mathcal{F}_{lc}(z_l) \frac{1 - \mathcal{G}_{cr}(\delta_l, r_l)^{n_l}}{1 - \mathcal{G}_{cr}(\delta_l, r_l)} = w_{cr,l} \ \forall\ l \in 1, 2, \cdots, L.$$

where f_i and g_i are convex functions of vector $\mathbf{x}$. We make use of the following two lemmas to transform (P2) in (19) to a DC optimization problem.

Lemma 1. *A function of product of two convex functions $f(x) \cdot g(x)$, $x \in \mathbb{R}$ is convex if $f(x)$ and $g(x)$ are positive and both the functions are either increasing or decreasing.*

Proof. Let us express $h(x) = f(x) \cdot g(x)$. Hessian h''_{xx} of $h(x)$ is:

$$h''_{xx} = f''_{xx} g(x) + f(x) g''_{xx} + 2 f'_x g'_x.$$

Hessian of a convex function is positive. Thus, f''_{xx} and g''_{xx} are > 0. As $f(x)$ and $g(x)$ are positive, the first two terms in h''_{xx} expression are positive. Given that both the functions are either increasing or decreasing, their derivatives are of same sign (either positive or negative). Thus, the last term is also positive. Consequently, h''_{xx} is positive and hence, $h(x)$ is convex. □

Lemma 2. *A function of product of two convex functions $f(x) \cdot g(y)$, $(x, y) \in \{\mathbb{R}^2 | f(x) + g(y) \geq 0\}$ can be written as a difference of two convex functions as:*

$$f(x) \cdot g(y) = \frac{(f(x) + g(y))^2}{2} - \left(\frac{f(x)^2}{2} + \frac{g(y)^2}{2} \right). \tag{21}$$

Proof. Using the binomial expansion of $(f(x) + g(y))^2$, we obtain the above expression for $f(x) \cdot g(y)$. From Lemma 1, square of a convex function is convex. Thus, $f(x)^2/2$ and $g(y)^2/2$ are convex. Consider $h(x, y) = (f(x) + g(y))^2$. Hessian $\mathbf{h}''$ of $h(x, y)$ is given as:

$$\begin{bmatrix} 2f''_{xx}(f(x) + g(y)) + 2(f'_x)^2 & 2f'_x g'_y \\ 2f'_x g'_y & 2g''_{yy}(f(x) + g(y)) + 2(g'_y)^2 \end{bmatrix}$$

Given that $f(x)$ and $g(y)$ are convex, diagonal terms in $\mathbf{h}''$ are positive. Determinant of $\mathbf{h}''$ is $4f''_{xx} g''_{yy}(f(x) + g(y)) + 4f''_{xx}(f(x)+g(y))(g'_y)^2 + 4g''_{yy}(f(x)+g(y))(f'_x)^2$. As $(x, y) \in \{\mathbb{R}^2 | f(x)+g(y) \geq 0\}$, determinant of $\mathbf{h}'' > 0$. Hence, $(f(x)+g(y))^2$ is also convex. □

Next, we make use of the two lemmas to reformulate (P2) to a DC optimization problem.

3) Optimization problem reformulation into DC: The number of RBs over LCN and number of slots over CRN used for the lth layer video transmission obtained from (14) and (15) respectively are nonconvex. Using epigraphs, these expressions are transformed into constraints as follows:

$$\frac{w_{lc,l}}{(z_l t_{rb} - \alpha)} \leq a_l \tag{22}$$

$$\frac{w_{cr,l}}{J(\delta_l)(r_l(t_{cr} - \delta_l) - \alpha)} \leq b_l. \tag{23}$$

By replacing the above expressions in (P2) with a_l and b_l, (P2), C2, C6a, and C7a are transformed to linear functions of a_l and b_l. These newly introduced constraints are converted to DC constraints as follows: We rewrite (22) as $w_{lc,l} - a_l z_l t_{rb} + a_l \alpha \leq 0$. The term $a_l z_l t_{rb}$ is nonconvex, which is converted to DC by using Lemma 2. Similarly, (23) is rewritten using (4) as $w_{cr,l} - b_l(p_{cr} p_m(\delta_l) + (1 - p_{cr})(1 - p_f(\delta_l)))(r_l(t_{cr} - \delta_l) - \alpha) \leq 0$. The terms $(1 - p_{cr})(1 - p_f(\delta_l)) r_l \delta_l$ and $b_l p_{cr} p_m(\delta_l)$ in the above expression are nonconvex and are converted to DC by Lemma 2.

The probability of transmission failure $\mathcal{F}_{lc}(z)$ over a LCN RB and the probability of transmission failure $\mathcal{F}_{cr}(r)$ due to channel fading and path loss over a slot in CRN from (1) and (5), respectively are nonconvex. Similarly, the expressions $\frac{1-\mathcal{F}_{lc}(z_l)^{n_l+1}}{1-\mathcal{F}_{lc}(z_l)}$, $\frac{1-\mathcal{F}_{lc}(z_l)^{n_l}}{1-\mathcal{F}_{lc}(z_l)}$, $\frac{1-\mathcal{G}_{cr}(\delta_l,r_l)^{n_l}}{1-\mathcal{G}_{cr}(\delta_l,r_l)}$, $\frac{1-\mathcal{G}_{cr}(\delta_l,r_l)^{n_l+1}}{1-\mathcal{G}_{cr}(\delta_l,r_l)}$, $\mathcal{F}_{lc}(z_l)^{n_l}$, and $\mathcal{G}_{cr}(\delta_l, r_l)^{n_l}$ appearing in the constraints C1a, C9, and C10 are also nonconvex. These are linear approximated using the first order Taylor series. For a function $f(x)$, its linear approximation is $f(x_\kappa) + \bigtriangledown f_x(x_\kappa)'(x - x_\kappa)$ around a point x_κ. Once these are approximated, the original constraints are expressed in DC using Lemmas 1 and 2.

Thus, the modified optimization problem constitutes a convex objective with DC constraints. Below, we use the convex-concave procedure (CCP) [39] to solve the DC optimization problem.

4) Solution to the DC optimization problem: In CCP, the DC functions $f_i(\mathbf{x}) - g_i(\mathbf{x})$ are approximated to convex functions. Linear approximation of the negative convex part, i.e., $g_i(\mathbf{x})$ is used to convert the DC to convex function. First order Taylor series is used to obtain the linear approximation. At a point $\mathbf{x}_k$ the linear approximation of the DC function is given as $f_i(\mathbf{x}) - g_i(\mathbf{x}) \approx f_i(\mathbf{x}) - g_i(\mathbf{x}_\kappa) - \bigtriangledown g_i(\mathbf{x}_\kappa)'(\mathbf{x} - \mathbf{x}_\kappa)$. Once the DC functions are convex approximated, convex optimization is applied to obtain the solution. At each iteration, once the optimal values are obtained, the approximation to the DC functions are improved to obtain a better solution. Convergence of this algorithm has been provided in [40].

A feasible initial point is required as input to this algorithm, which may be challenging given the number of variables

Algorithm 2: Penalty convex-concave procedure.

Initial point $\mathbf{x}_0$, $\tau_0 > 0$, τ_{max}, μ, and ς. $\kappa = 0$

while $|\Gamma_\kappa - \Gamma_{\kappa-1}| < \varsigma$ **do**

1. Form $\hat{g}_j(\mathbf{x}; \mathbf{x}_\kappa) \triangleq g_j(\mathbf{x}_\kappa) + \bigtriangledown g_j(\mathbf{x}_\kappa)'(\mathbf{x} - \mathbf{x}_\kappa)$ for $j = 0, 1, \cdots, M$.

2. Set the value of $\mathbf{x}_{\kappa+1}$ to the solution of

$$\begin{aligned} \Gamma_\kappa = \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) - \hat{g}_0(\mathbf{x}, \mathbf{x}_\kappa) + \tau_\kappa \sum_{j=1}^{M} v_j \\ \text{subject to} \quad & f_j(\mathbf{x}) - \hat{g}_j(\mathbf{x}; \mathbf{x}_\kappa) \leq v_j, \forall\ j \in 1, \cdot, M, \\ & v_j \geq 0, j = 1, 2, \cdots, M. \end{aligned}$$

3. Update τ: $\tau_{\kappa+1} = min(\mu\tau_\kappa, \tau_{max})$.

end

involved. In [41], it was shown that the need for an initial feasible point could be overcome by making use of slack variables in the constraints and penalizing the violations. Naming this modified algorithm as penalty CCP, Algorithm 2 describes the steps involved. In step 1, the DC constraints are linear approximated around point $\mathbf{x}_\kappa$. Slack variable v_j is considered for each constraint, τ is the penalty factor added to the objective, and μ is the multiplying factor which increments the penalty factor in each iteration. τ_{max} is the upper limit of τ. In each iteration, an approximate convex optimization problem is solved in step 2. The algorithm stops when the change in the objective function value is less than a small constant ς.

5) Complexity analysis and convex problem decomposition: In step 2 of the Algorithm 2, a convex optimization problem is solved. For solving (P2), number of variables are $(20L+4)$, where $7L$ are the required parameters (z_l, δ_l, r_l, u_l, n_l, $p_{lc,l}^{rtx}$, and $p_{cr,l}^{rtx}$), $4L$ are the additional introduced variables ($w_{lc,l}$, $w_{cr,l}$, a_l, and b_l), and $9L + 4$ are slack variables (v_i) corresponding to each constraint. Solving this convex problem involves high complexity. Therefore, we resort to the decomposition method to solve this convex problem.

To reduce the computational complexity, dual decomposition of the approximated convex optimization problem is carried out. Due to the specific structure of the problem, it is decomposed with each layer forming a separate subproblem. The Lagrange function of the approximated convex problem is decomposed into L subproblems corresponding to each layer and $9L+4$ subproblems corresponding to the slack variables. The subproblems corresponding to each layer are nonlinear, while the subproblems corresponding to the slack variables are linear. The problem is iteratively solved, where the subproblems are fed with the Lagrange multipliers to obtain the optimal parameters. Using these optimal values from each subproblem, the main problem computes the optimal Lagrange multipliers using the subgradient method [42].

This decomposition results in a total of $10L+4$ subproblems with 15 variables for the L subproblems corresponding to each layer and 1 variable for the $9L+4$ subproblems corresponding to each slack variable. Complexity for solving subproblems corresponding to L layers is $\mathcal{O}(15^3)$ and they can be solved in parallel. Closed form expressions are obtained for solving the $9L+4$ subproblems corresponding to each slack variable. Their computation complexity is negligible. The computation of optimal Lagrange multipliers requires only a few arithmetic operations. Thus, the decomposition method reduces the complexity significantly.

We performed the optimization in Matlab R2014a running on Intel i7-3770 CPU with 3.4 GHz clock and 16 GB RAM. On an average, the penalty CCP along with the decomposition algorithm took around 6 minutes to converge to the optimal solution for 4-layer video transmission.

For CM-basic, the parameters in (P2) are made layer independent. Thus, decomposition method is not required in CM-basic. The solution to (P2) for CM-basic is obtained using the penalty CCP method as proposed in Algorithm 2. There are 22 variables, where 7 are the desired parameters, 4 are the additional variables introduced, and 13 are the slack variables corresponding to the constraints. Complexity of solving this problem is $\mathcal{O}(22^3)$. Though its complexity is lower than CM-opt, as will be noted in Section V, the performance of CM-basic is inferior to CM-opt.

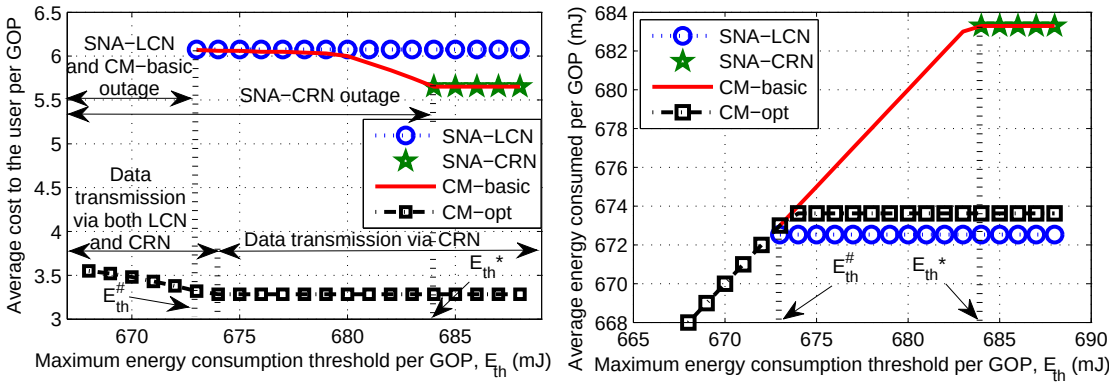
6) Discussion: The optimization problem (P2) provides optimal parameters for the transmission of L-layer video content via the two networks. The user-provided parameters (namely, $E_{th}^{(k)}$, $Q_{th}^{(k)}$, and $L^{(k)}$) change with the remaining battery level of the device, which is on the order of minutes. While these computations are done on the BS, which have high processing capabilities, the optimization problem computation can be easily parallelized to meet the deadlines.

V. System Simulation Results

In this section we evaluate the user cost in video reception via optimized CM (CM-opt) and compare it with CM-basic, SNA-LCN, and SNA-CRN (cf. Section III-E). A single cell scenario is considered with 1 Km cell radius. LCN RBs are of duration 0.5 ms with bandwidth of 5 MHz (LTE in 900 MHz band is considered). CRN operation is considered over UHF band (500 MHz) with PU channel bandwidth of 6 MHz. Slot duration is 10 ms (as in IEEE 802.22). Channel idling probability is 0.7 [42]. Misdetection probability threshold is set to 0.05 [34]. Transmission power over LCN and CRN is 24 dBm and 20 dBm, respectively, while noise variance is 10^{-4} [43]. Path loss exponent is 3.76 [44].

$\phi_{cr}/\phi_{lc} \triangleq \rho$ is the ratio of cost charged by CRN and LCN per unit time. In this study we consider that the user is charged monetary cost 100 units per unit time for operation over LCN, and the default value of cost ratio $\rho = 0.5$ (50 units per unit time for PU spectrum leasing). Device power consumption for reception is 1737 mW, while its tail power consumption is 1325 mW [37]. Device energy consumption is a function of the number of layers it intends to receive. In 3GPP, channel state feedback is provided within 6 ms [45]. Hence, for each packet reception error, the receiver can inform the BS almost within the next CRN slot. For this study we consider maximum retry limit $n_{max} = 5$.

Foreman and *Mobile* QCIF video sequences are encoded into SNR scalable bit-streams with a base layer and 3 enhancement layers, with 30 frames per second (fps) and with GOP of



(a) (b)

Fig. 2: Variation in (a) user cost and (b) energy consumed, versus energy consumption threshold. $Q_{th} = 95\%$. $\rho = 0.5$, $p_{cr} = 0.7$, $L = 4$.

TABLE II: Optimized parameters for CM-opt. $E_{th} = 670$ mJ and $p_{cr} = 0.7$.

Parameter	**Layer 1**	**Layer 2**	**Layer 3**	**Layer 4**
Transmission rates over LCN (Mbps)	10.17	10.99	11.91	41.6
Transmission rates over CRN (Mbps)	11.4	12.8	18.3	48.0
Sensing time over CRN (ms)	1.88	1.80	1.58	1.4
Fraction of data transmitted via CRN	0	0	0.2	1
Average number of retry attempts	3.9	3.5	0.9	0.3
Probability of retry over LCN for data originally transmitted over LCN	1	1	0.74	0.44
Probability of retry over CRN for data originally transmitted over CRN	0	0	0	0.9

duration 16 frames having structure "IPPPPPPPPPPPPPP". For the *Foreman* sequence, base layer data rate is 69 Kbps, while the enhancement layers 2, 3, and 4 are encoded at rates 58, 95, and 239 Kbps, respectively. For the *Mobile* sequence, base layer data rate is 117 Kbps, while the enhancement layers 2, 3, and 4 are encoded at rates 80, 174, and 656 Kbps, respectively. The encoding is done using JSVM software [46]. This data is suitably divided into packets with per-packet overhead of 24 Bytes, as in IEEE 802.11.

A. Video Transmission to a Single CM User

A user located at a distance of 0.5 Km from the BS requests for *Foreman* video sequence with 4 layers. The minimum desirable video quality is 95%. In the following, we study the single-user performance in terms of device and network constraints and user preference.

1) Effects of user preference: Figs. 2(a) and 2(b) present respectively the user cost and energy consumed versus energy consumption threshold. Compared to SNA-LCN, SNA-CRN offers service at a lower cost due to its opportunistic spectrum access. However, operation over CRN requires a higher minimum energy E^*_{th} as opposed to $E^{\#}_{th}$ in SNA-LCN (i.e., $E^*_{th} \geq E^{\#}_{th}$, as noted in Fig. 2(b)), because of the additional activity of the device due to intermittent sensing and channel imperfections. CM-basic provides service in all regimes as it exploits both CRN and LCN. For $E_{th} > E^*_{th}$, cost is same as in SNA-CRN, because all content is transmitted via CRN. As the device energy consumption constraint becomes more stringent ($E_{th} < E^{\#}_{th}$), the SNA-LCN and CM-basic go into outage because energy consumption in SNA-LCN exceeds the threshold. In contrast, less resource requirement in CM-opt allows to serve a user with even more stringent energy constraints. Though, the energy consumption in CM-opt is higher than that in SNA-LCN, CM-opt serves the user at a lower cost due to concurrent transmissions via LCN and CRN. Overall, CM-opt outperforms CM-basic (hence SNA-LCN and SNA-CRN) by reducing the cost on average by up to 44.1%.

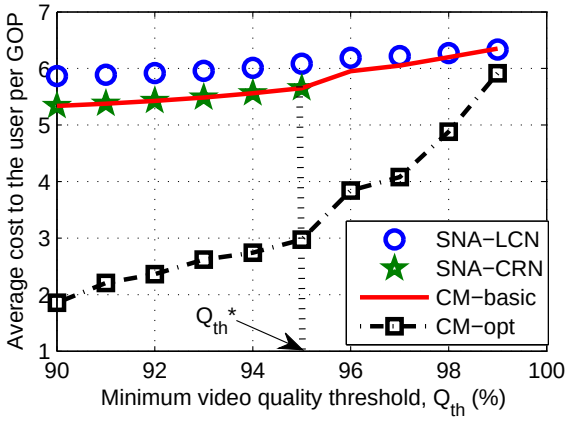
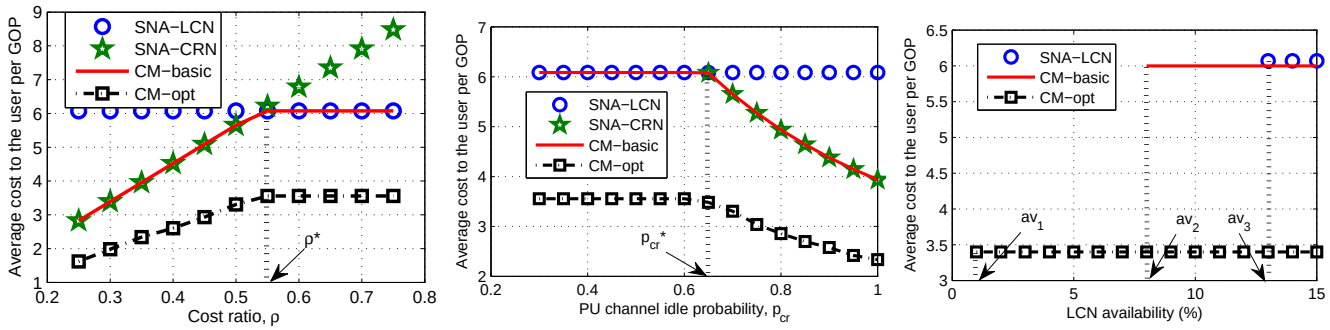


Fig. 3: Variation in user cost versus video quality constraint. $E_{th} = 685$ mJ, $\rho = 0.5$, $p_{cr} = 0.7$, $L = 4$.

Table II presents the values of different layer-dependent parameters that are optimized with $E_{th} = 670$ mJ. As anticipated, to protect a higher priority layer from channel-induced packet losses, its data rate is low. Also, the sensing duration and number of retransmissions are higher for high priority layers, and most of the retransmissions are done via LCN for higher priority layers.

Fig. 3 presents the impact of video quality threshold Q_{th}



(a) (b) (c)

Fig. 4: User cost versus change in (a) cost ratio ρ, with $p_{cr} = 0.7$; (b) PU channel idling probability p_{cr}, with $\rho = 0.5$; (c) LCN resource availability with $p_{cr} = 0.7$, $\rho = 0.5$. $Q_{th} = 95\%$, $E_{th} = 685$ mJ, and $L = 4$.

on user cost. Q_{th} is varied from 90% to 99%, with $E_{th} = 685$ mJ. A higher Q_{th} requires higher resources, thereby increasing energy consumption and cost. With the considered E_{th}, SNA-CRN can guarantee service only up to $Q_{th} \leq Q^*_{th}$. In contrast, SNA-LCN meets the user demand at a lower energy consumption, though at a higher cost. CM performs better by using CRN at lower Q_{th} for reduced cost, and optimally dividing the stream over LCN and CRN to meet higher Q_{th} at a reduced cost.

Remark 1: Compared to SNA, CM-opt provides on an average 42.7% cost reduction along with providing service to the users with more stringent E_{th} and Q_{th} requirements.

2) Effects of network parameters: Fig. 4(a) shows the cost to the user in various schemes with the change in CRN versus LCN cost ratio ρ. Observe that cost to the user increases with the increase in ρ (ϕ_{cr}). User cost in SNA-CRN surpasses SNA-LCN at $\rho = \rho^*$. To achieve a low-cost transmission, the proposed CM-opt delivers all the data via CRN for $\rho < \rho^*$, while the data is transmitted via LCN for $\rho > \rho^*$.

The effect of PU activity parameter on user cost is plotted in Fig. 4(b). PU channel idling probability p_{cr} is varied from 0.25 to 1. Lower idling probability induces the use of higher number of CRN slots to transmit a video content, thus increasing energy consumption and cost. For $p_{cr} < p^*_{cr}$, the energy consumption constraint is violated in SNA-CRN resulting in service outage. CM-opt chooses the radio access optimally to maintain low cost to the user at all times.

High demands from other users may push the single network operation into outage. As an example, we consider that, for a single user operation 5% resource is available via CRN, and LCN resource availability is varied from 1% to 15%. Given the user's resource request, it cannot be served via SNA-CRN. Fig. 4(c) shows that, SNA-LCN is able to fulfill the user requirement only above 13% resource availability (marked av_3). CM provides service by simultaneously using both LCN and CRN resources. CM-basic is useful only when the available LCN resources are more than 8% (marked av_2). Layer based parameter adaptation in CM-opt requires less resources. Hence, it provides service at even lower LCN resource availability (see av_1 and av_2).

3) Effects of user device battery energy constraint: As discussed in Section III-C, to aid continued reception, user's requested video quality and maximum energy consumption constraint is proposed to be a function of its remaining battery level. For demonstration, we consider a mobile device with maximum energy consumption threshold and requested number of video layers L, as shown in Fig. 5.

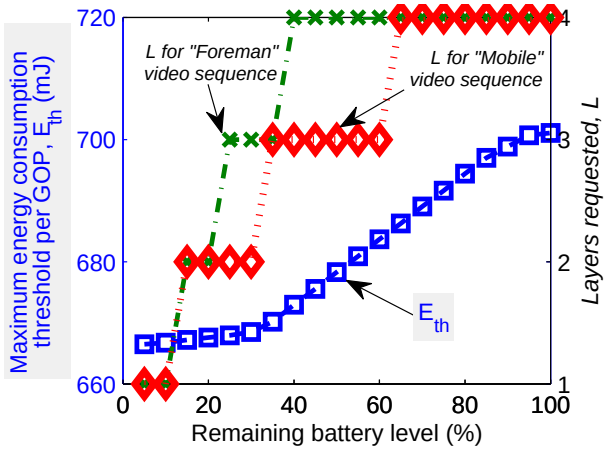
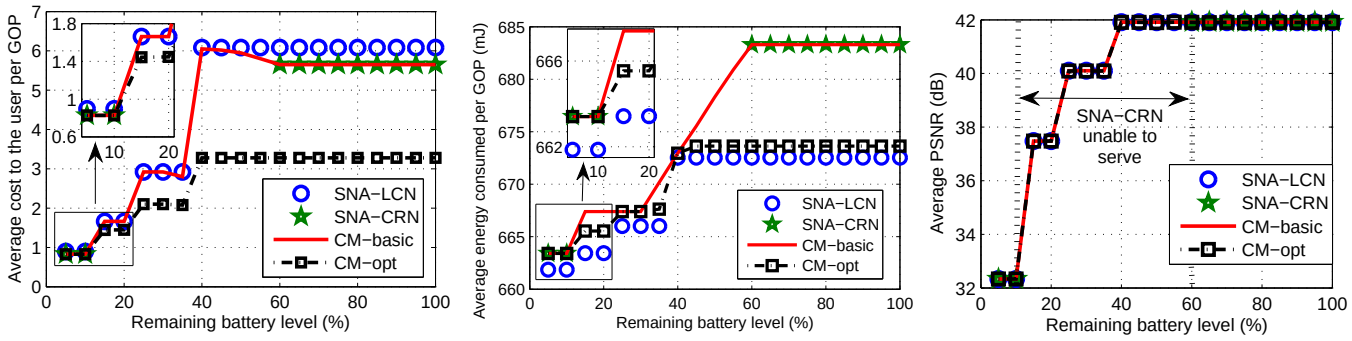


Fig. 5: User's demanded video quality and energy constraint with the remaining battery level.

A lower remaining battery level induces lower video quality and E_{th}. Q_{th} is maintained at 95%. Fig. 6(a) presents the user cost in video transmission in the various schemes. SNA-CRN goes into outage for some E_{th} regimes, though providing service at a lower cost. The gain in CM-opt as compared to SNA as well as CM-basic scheme is higher when L is higher. This is because, the advantage with layer-dependent parameter adaptation diminishes for lower number of requested layers.

Figs. 6(b) and 6(c) present respectively the user energy consumption and mean received video PSNR for different battery levels of the user device. Energy consumed in all the cases remain below the maximum energy consumption constraint, while the PSNR increases with the increased number of layers requested. As Q_{th} is maintained at 95%, the PSNR remains constant for a fixed number of layers requested.

Remark 2: CM-opt intelligently distributes data via the two networks depending on the individual network cost and availability, to ensure low cost to the user without affecting the video reception quality.



(a) (b) (c)

Fig. 6: User (a) cost, (b) energy consumption, and (c) received video quality (PSNR) with the remaining battery level for the *'Foreman'* video request. $Q_{th} = 95\%$, $\rho = 0.5$, and $p_{cr} = 0.7$.

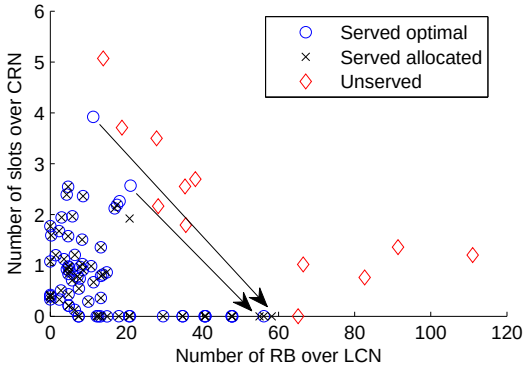


Fig. 7: Requested and allocated resources to the users via the two networks for a network instance in Scenario 3. $\mathbb{N} = 80$, $Q_{th} = 95\%$, $\rho = 0.5$, and $p_{cr} = 0.7$.

B. Multiuser Operation

We now consider a single cell scenario with 80 CM users uniformly randomly distributed across the cell. We consider three network scenarios. In the first, all users request *'Foreman'* video sequence, while in the second all users request *'Mobile'* video sequence. In the third scenario, the users request one of the two video sequences: *'Foreman'* or *'Mobile'*, with equal probability. Remaining battery level of user device is distributed uniformly between $[0\ 100]\%$. Depending on the remaining battery level, the user's maximum energy consumption constraint and number of layers requested are obtained from the characteristics in Fig. 5. The minimum video quality constraint is fixed to 95% for all users. We consider three random instances of the network in each of the three scenarios and evaluate the average performance.

Fig. 7 presents an example of resources requested/allocated via the two networks for a random network instance in the third scenario. Optimal resources requested by the individual users are computed using (P2) in (19). Following Algorithm 1, the users are served in an increasing order of their cost, or equivalently the number of layers requested. The users meeting the resource availability criteria are served (marked *allocated* in Fig. 7). If the CRN resources are insufficient, (P2) is recomputed to obtain the revised resource allocation. If a user's revised resource allocation is acceptable, the user is served (marked by arrow in Fig. 7). A user is dropped if its resource requirement is not met. Overall, the users requesting lower resources are served, while the users with higher demands are more likely to be dropped.

Figs. 8(a) and 8(b) presents respectively the total network cost $\mathcal{C}^*_{net}$ (cf. (P1) in (18)) and the average number of users served in the three scenarios. Number of users served is higher in scenario 1 as the data rate of the *Foreman* video sequence is lower compared to the *Mobile* video sequence. Total network cost in SNA schemes is seen to be higher than (inferior to) the CM system because the users are optimally allocated resources from the two networks simultaneously in CM. Further, the number of users served in CM is higher compared to the other schemes. Fig. 8(c) shows the average PSNR of the users served. We observe that the average PSNR in CM-opt is better than that in all the other schemes as higher number of users are served in CM-opt. Due to layer-dependent parameter optimization, CM-opt outperforms SNA-LCN schemes by serving on average 48.2% more users with about 3% higher PSNR on average.

To gain further insights, Figs. 9(a) and 9(b) presents the average cost per user and fraction of users served versus number of layers requested in the first scenario. From Fig. 9(a) it is observed that the average cost per user served is the lowest in CM-opt, which is due to its layer-dependent parameter optimization. Cost reduction in CM-opt is higher for higher number of requested layers as noted in Section V-A3. As the users are served in an increasing order of the number of layers requested, the users requiring higher number of layers are more likely to be dropped (see Fig. 9(b)). However, in CM system, where the data is optimally transmitted via the two networks, the fraction of users served is higher. Similar trends are observed in the other scenarios as well.

Remark 3: Compared to SNA, CM system serves a higher number of users with a higher average video reception quality and at a lower cost.

VI. Conclusion

In this paper, a novel paradigm, called cognitive multi-homing (CM), has been presented, where the user data is

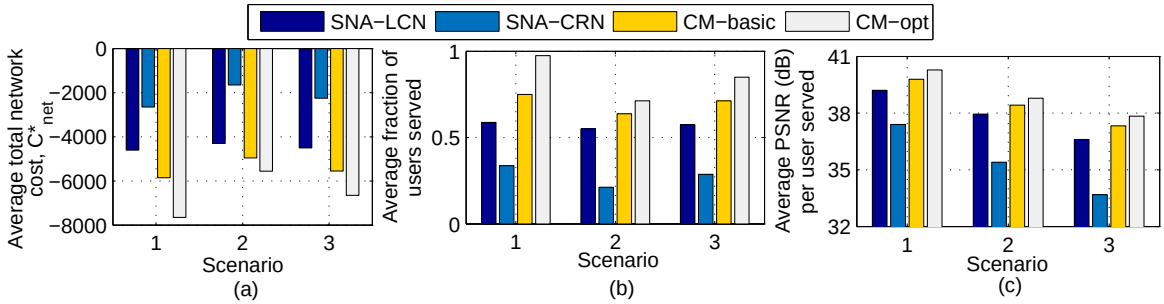


Fig. 8: Average (a) total network cost, (b) fraction of users served, and (c) PSNR per user served in the three scenarios. $\mathbb{N} = 80$, $Q_{th} = 95\%$, $\rho = 0.5$, $p_{cr} = 0.7$, and $\beta = 100$.

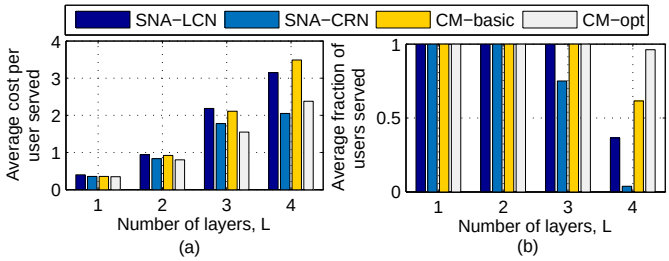


Fig. 9: (a) Average cost per user served and (b) average fraction of users served for users requesting different number of layers in Scenario 1. $\mathbb{N} = 80$, $Q_{th} = 95\%$, $\rho = 0.5$, $p_{cr} = 0.7$, and $\beta = 100$.

optimally split across LCN and CRN before transmission to the multihomed users. The CM approach aids in mitigating the spectrum scarcity problem of LCN, which CRN alone cannot solve. In the proposed approach, dynamic radio resource selection, sensing duration and transmission rate adaptation over CRN, and transmission rate adaptation over LCN are optimized, which reduces the cost to the user while meeting the energy consumption and video quality thresholds. The optimized CM has been shown to outperform the SNA schemes by reducing the cost up to 44%, offering about 3% high PSNR in video reception quality, and serving almost 48% more number of users. CM can be easily implemented on devices with multihoming and CR capabilities.

ACKNOWLEDGMENTS

This work has been supported in parts by the ITRA Media Lab Asia project under Grant no. ITRA/15(63)/Mobile/MBSSCRN/01 and the Department of Science and Technology under Grant SB/S3/EECE/0248/2014.

REFERENCES

[1] C. Cisco Systems, San Jose. (Feb. 2016) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2015–2020 White Paper.

[2] C. Xu, T. Liu, J. Guan, H. Zhang, and G.-M. Muntean, "CMT-QA: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2193–2205, Nov. 2013.

[3] M. Lopez-Benitez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2091–2104, Jun. 2013.

[4] N. Freris, C.-H. Hsu, J. Singh, and X. Zhu, "Distortion-aware scalable video streaming to multinetwork clients," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 469–481, Apr. 2013.

[5] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 688–701, Apr. 2015.

[6] O. C. Kwon, Y. Go, Y. Park, and H. Song, "MPMTP: Multipath multimedia transport protocol using systematic raptor codes over wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 9, pp. 1903–1916, Sep. 2015.

[7] J. Wu, C. Yuen, M. Wang, and J. Chen, "Content-aware concurrent multipath transfer for high-definition video streaming over heterogeneous wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 3, pp. 710–723, Mar. 2016.

[8] S.-P. Chuah, Z. Chen, and Y.-P. Tan, "Energy minimization for wireless video transmissions with deadline and reliability constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 3, pp. 467–481, Mar. 2013.

[9] W.-H. Kuo, R. Kaliski, and H.-Y. Wei, "A QoE-based link adaptation scheme for H.264/SVC video multicast over IEEE 802.11," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 812–826, May 2015.

[10] C. Chen, R. Heath, A. Bovik, and G. de Veciana, "A Markov decision model for adaptive scheduling of stored scalable videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1081–1095, Jun. 2013.

[11] F. Fu and M. van der Schaar, "Structural solutions for dynamic scheduling in wireless multimedia transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 727–739, May 2012.

[12] Y. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. Alnuweiri, "A link adaptation scheme for efficient transmission of H.264 scalable video over multirate WLANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 875–887, Jul. 2008.

[13] H. Abou-zeid, H. Hassanein, and S. Valentin, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2013–2026, Jun. 2014.

[14] W. Hu and G. Cao, "Energy-aware video streaming on smartphones," in *Proc. IEEE INFOCOM*, Hong Kong, Apr. 2015, pp. 1185–1193.

[15] J. Wu, C. Yuen, N. M. Cheung, J. Chen, and C. W. Chen, "Modeling and optimization of high frame rate video transmission over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2713–2726, Apr. 2016.

[16] D. Hu, S. Mao, Y. Hou, and J. Reed, "Scalable video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 334–344, Apr. 2010.

[17] X.-L. Huang, G. Wang, F. Hu, and S. Kumar, "The impact of spectrum sensing frequency and packet-loading scheme on multimedia transmission over cognitive radio networks," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 748–761, Aug. 2011.

[18] K. Tan, K. Kim, Y. Xin, S. Rangarajan, and P. Mohapatra, "RECOG: A sensing-based cognitive radio system with real-time application support," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2504–2516, Nov. 2013.

[19] Q. Jiang, V. Leung, M. Pourazad, H. Tang, and H. Xi, "Energy-efficient adaptive transmission of scalable video streaming in cognitive radio communications," *IEEE Syst. J.*, vol. 10, no. 2, pp. 761–772, Jun. 2016.

[20] R. Yao, Y. Liu, J. Liu, P. Zhao, and S. Ci, "Utility-based H.264/SVC video streaming over multi-channel cognitive radio networks," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 434–449, Mar. 2015.

[21] X. Chen and C. Yuen, "Efficient resource allocation in a rateless-coded MU-MIMO cognitive radio network with QoS provisioning and limited feedback," *IEEE Trans. Vehicular Technol.*, vol. 62, no. 1, pp. 395–399, Jan. 2013.

[22] T. Jiang, H. Wang, and A. V. Vasilakos, "QoE-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1215–1224, Aug. 2012.

[23] Z. He, S. Mao, and S. Kompella, "A decomposition approach to quality-driven multiuser video streaming in cellular cognitive radio networks," *IEEE Trans Wireless Commun.*, vol. 15, no. 1, pp. 728–739, Jan. 2016.

[24] W. Song and W. Zhuang, "Performance analysis of probabilistic multipath transmission of video streaming traffic over multi-radio wireless devices," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1554–1564, Apr. 2012.

[25] J. Wu, C. Yuen, N.-M. Cheung, and J. Chen, "Delay-constrained high definition video transmission in heterogeneous wireless networks with multi-homed terminals," *IEEE Trans. Mobile Comput.*, vol. 15, no. 3, pp. 641–655, Mar. 2016.

[26] M. Ismail and W. Zhuang, "Mobile terminal energy management for sustainable multi-homing video transmission," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4616–4627, Aug. 2014.

[27] J. Lee and S. Bahk, "On the MDP-based cost minimization for video-on-demand services in a heterogeneous wireless network with multihomed terminals," *IEEE Trans. Mobile Comput.*, vol. 12, no. 9, pp. 1737–1749, Sep. 2013.

[28] A. H. Zahran and C. Sreenan, "Threshold-based media streaming optimization for heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 6, pp. 753–764, Jun. 2010.

[29] S. Agarwal, S. De, S. Kumar, and H. Gupta, "QoS-aware downlink cooperation for cell-edge and handoff users," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2512–2527, Jun. 2015.

[30] K. Habak, K. A. Harras, and M. Youssef, "Bandwidth aggregation techniques in heterogeneous multi-homed devices," *Comput. Netw.*, vol. 92, no. P1, pp. 168–188, Dec. 2015.

[31] Y. Go, O. C. Kwon, and H. Song, "An energy-efficient HTTP adaptive video streaming with networking cost constraint over heterogeneous wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1646–1657, Sep. 2015.

[32] R. Zhang, M. Wang, L. Cai, Z. Zheng, and X. Shen, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, Jun. 2015.

[33] S. Agarwal, S. De, and J.-B. Seo, "Cognitive multihoming: Maximizing network utility over CR-assisted cellular network," in *Proc. IEEE SDRANCAN GLOBECOM Workshop*, San Diego, CA, USA, Dec. 2015.

[34] Y. Pei, Y.-C. Liang, K. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1648–1659, Sep. 2011.

[35] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[36] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for Internet video," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 339–350, Aug. 2013.

[37] S. Deng and H. Balakrishnan, "Traffic-aware techniques to reduce 3G/LTE wireless energy consumption," in *Proc. Int. Conf. Emerging Netw. Experiments and Technol.*, Nice, France, Dec. 2012, pp. 181–192.

[38] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: A measurement study and implications for network applications," in *Proc. ACM SIGCOMM Conf. Internet Measurement*, Chicago, IL, USA, Nov. 2009, pp. 280–293.

[39] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[40] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," in *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc., 2009, pp. 1759–1767.

[41] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, 2016.

[42] S. Geirhofer, L. Tong, and B. Sadler, "Cognitive radios for dynamic spectrum access - Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.

[43] F. Zhao, L. Wei, and H. Chen, "Optimal time allocation for wireless information and power transfer in wireless powered communication systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1830–1835, Mar. 2016.

[44] Y. Long, H. Li, H. Yue, M. Pan, and Y. Fang, "SUM: Spectrum utilization maximization in energy-constrained cooperative cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2105–2116, Nov. 2014.

[45] 3GPP, "Further advancements for E-UTRA physical layer aspects," *3rd Generation Partnership Project, TR 36.814*, 2010.

[46] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-12," in *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG Doc. JVT-Y202*, Oct. 2007.

Satyam Agarwal (S'13) received his B.Tech. in Electronics and Communication from Thapar University, India, in 2010 and M.Tech. in Electrical Engineering from IIT Kanpur in 2012. He is currently working towards the Ph.D. degree in the Department of Electrical Engineering at IIT Delhi. His research interests include cooperative communications and link layer protocol designs in wireless networks. He is a student member of IEEE and IEEE Communications Society.

Swades De (S'02-M'04-SM'14) received the Ph.D. degree from the State University of New York at Buffalo, NY, USA, in 2004. In 2004, he worked as an ERCIM researcher at ISTI-CNR, Italy. From 2004 to 2006 he was with New Jersey Institute of Technology, NJ, USA, as a tenure-track Assistant Professor. He is currently an Associate Professor in the Department of Electrical Engineering, IIT Delhi, India. His research interests include performance study, resource efficiency in wireless networks, broadband wireless access, and optical communication systems.