LTE Multimedia Broadcast Multicast Service Provisioning Based on Robust Header Compression

Chen Jiang, *Student Member, IEEE*, Wenhao Wu, *Student Member, IEEE*, and Zhi Ding, *Fellow, IEEE*

***Abstract*— One important issue that confronts network service providers is the need to provide reliable multimedia data service efficiently over cellular networks for a large number of subscribers under dynamic channel conditions. In long term evolution (LTE) networks, multimedia broadcast multicast service (MBMS) is a bandwidth efficient data service to simultaneously support multiple users at high bandwidth efficiency. In this paper, instead of considering spectrum resource allocation, we investigate MBMS provisioning for each mobile user based on the higher layer robust header compression (ROHC) consideration in response to user channel quality to reduce packet losses. We formulate a profit maximization problem for two different MBMS channel models and further propose a new MBMS assignment scheme for each user to be assigned a target MBMS with optimal ROHC parameters. We further develop a dynamic programming algorithm for user assignment and ROHC parameters optimization to achieve maximal profit with high spectrum resource utility. Our numerical results demonstrate substantial profit gain achieved by the proposed method in LTE systems.**

***Index Terms*— Multimedia broadcast multicast service, robust header compression, hidden Markov model, dynamic programming.**

I. Introduction

MANY of today's cellular subscribers are smartphone users with frequent access to multimedia contents. Since the primary cost for a network service provider (NSP) is its bandwidth usage, it is therefore natural for NSPs to maximize their profit by providing multimedia services of satisfactory quality to subscribers for achieving high spectral efficiency. According to a large number of users and user diversity, the allocation of insufficient bandwidth is always a big challenge for the NSP to maintain a higher and stable quality of wireless transmission due to the huge demand of multimedia. In this work, we use the NSP profit as an objective to investigate the tradeoff between bandwidth utility and multiple-user satisfaction of multimedia transmission in cellular wireless networks.

To effectively deliver volumes of multimedia data service to a large number of users at same time, Multimedia Broadcast Multicast Service (MBMS) provides an efficient interface. Instead of always relying on individual cellular links, MBMS offers a highly efficient service which includes both point-to-point (P-t-P) radio bearers and point-to-multipoint (P-t-M) radio bearers in the 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) [1]. Traditional unicast (P-t-P) assigns a dedicated radio bearer to every active user according to the specific channel condition. Using one unicast (P-t-P) bearer for each of many subscribers requires large service bandwidth to satisfactorily deliver the media contents with Quality of Service (QoS) required by the users. On the other hand, one multicast (P-t-M) radio bearer can simultaneously serve multiple subscribers interested in acquiring the same popular data content, thereby significantly improving both spectrum efficiency and transmit power efficiency [2]. Since it is difficult to guarantee the QoS of all subscribers by establishing only a single P-t-M radio bearer, Multiple MBMS radio bearers can be set up to serve users of different channel conditions. Thus, MBMS radio bearer provisioning between P-t-P and P-t-M is an important decision for improving NSP profit. The tradeoff between collecting higher revenue from users, which is directly associated with user QoS, and reducing the cost of bandwidth usage associated with the number of MBMS bearers is an important design consideration to be tackled in this work.

Currently, most existing works of radio bearer selection for MBMS rely on some user counting mechanisms which use either P-t-P or P-t-M exclusively but not both [3]. The typical performance objectives are power efficiency, throughput and transmission performance. The authors of [4]–[6] studied how to improve the multicast performance in LTE networks. The improvement of power efficiency by selecting different transmission methods is studied in [7] and [8]. A new framework proposed by [9] attempted to maximize the profit based on bandwidth allocation for the wireless video broadcasting system. All these existing works assume using either n P-t-P bearers or one single P-t-M bearer to serve n subscribers. Such limitation apparently is not flexible enough to achieve an optimal tradeoff between spectrum efficiency and bandwidth usage when the subscribers have diverse channel conditions. Instead of limiting the NSP to either pure

Manuscript received May 12, 2017; revised September 2, 2017 and November 8, 2017; accepted November 11, 2017. Date of publication November 28, 2017; date of current version February 9, 2018. This work was supported by the National Science Foundation under Grant 1321143 and Grant 1443870. This paper was presented at the IEEE WCNC Spring, San Francisco, CA, USA, March 2017. The associate editor coordinating the review of this paper and approving it for publication was D. W. K. Ng. *(Corresponding author: Chen Jiang.)*

The authors are with the Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616 USA (e-mail: cxjiang@ucdavis.edu; wnhwu@ucdavis.edu; zding@ucdavis.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/TWC.2017.2776906

unicast (P-t-P) or pure multicast (P-t-M), we propose to study the tradeoff between user QoS revenue and bandwidth usage cost through adaptive user assignment to various MBMS bearers, each of which serves a subset of subscribers whose channel qualities support similar QoS requirements. A related work [10] proposed a hybrid transmission approach to find the optimum configuration which adjusts the system to either streaming or file delivery service by considering MBMS Single Frequency Network (MBSFN) area size and Evolved Multimedia Broadcast Multicast Services (E-MBMS) data rate.

One less known yet indispensable mechanism that plays an important role in many MBMS applications is RObust Header Compression (ROHC). Located in the packet data convergence protocol (PDCP) sub-layer of Layer 2 (L2) of the LTE protocol stack [11], ROHC is a key interface between cellular and Internet Protocol (IP) networks and can substantially improve cellular network's bandwidth efficiency by reducing typical IP headers from as long as 40 bytes (e.g. IPv4/TCP) to as short as 1 byte based on exploring the correlation between header fields from the consecutive packets of the same stream. For many applications such as Voice over Internet Protocol (VoIP), interactive gaming and video streaming, etc., the size of the payload in each packet can be comparable to or even smaller than that of the headers if un-compressed, leading to poor transmission efficiency. ROHC is responsible for the header compression in wireless networks with poorer channel conditions in order to achieve a trade-off between reliability and efficiency [12]. Surprisingly, although ROHC has been standardized and widely incorporated in many protocol stacks, it has not been widely studied among traditional wireless research publications and many questions remain unanswered.

There exist several works that mainly focus on the performance analysis of ROHC itself in specific networks. The compression efficiency of ROHC over mobile WiMAX is studied in [13]. In [14], two ROHC implementations known as ROHCv1 and ROHCv2 are evaluated in terms of potential throughput gain and complexity for wireless IP networks in multimedia delivery. Reference [15] presented an efficient hardware architecture for accelerating ROHCv2 based on LTE energy efficiency. In [16], the authors investigated the resource consumption and potential performance gain of implementing ROHC encoding and decoding functions and studied the architectural implications of ROHC in future networks. The authors of [17] investigated the impact of a Window-based Least Significant Bits (W-LSB) encoding on the performance of ROHC. In [18], the video streaming file transmission over MBMS studied the use of ROHC to generate a single RTP/UDP/IP packet flow in Packet Data Convergence Protocol (PDCP). However, existing works on ROHC have focused only on one single packet-switched link. There is a lack of understanding on the control of ROHC in MBMS provisioning, in which the proper use of ROHC can improve transmission efficiency in many MBMS applications.

There are two major challenges in optimizing the ROHC settings for MBMS. Firstly, as with ROHC for an individual wireless link, inappropriate settings of ROHC parameters may cause frequent decompression failure by the ROHC decompressor at the receiver end, thereby resulting in serious loss of data packets even if they are correctly received. In our initial work [19], we have investigated the problem of ROHC parameter optimization for a single user and demonstrated substantial improvement of transmission efficiency. We also developed an adaptive ROHC scheme under limited radio resources. Moreover, for MBMS (P-t-M), there is one single ROHC entity to serve a group of users. We note that, when establishing MBMS bearers, ROHC parameters must also be optimized for each service group of users whose channel conditions tend to differ. This design requirement becomes even more challenging for our proposed novel MBMS schemes with adaptive user assignment to various MBMS bearers. In this new scenario, the NSP must jointly design the MBMS user assignment when establishing MBMS service groups and optimize the ROHC parameters for each individual group in which users experience diverse channel qualities.

In this work, we study the problem of NSP profit maximization in MBMS through dynamic user assignment and ROHC optimization. We focus on optimizing MBMS user grouping in accordance with each MBMS user's ROHC control parameter which can be optimized for maximum transmission efficiency. We establish a new system based on a more practical channel model beyond our preliminary works [20]. Based on this more advanced and practical system model, we develop an optimized MBMS service provisioning algorithm in response to dynamic user channel conditions. Our major contributions are as follows:

- We fully consider the diversity of channel qualities among active users for MBMS packet services. Two channel models are applied to characterize such diversity. In the first model, the packet channel is derived from the PHY fading channel into a Markov model for each user. In the second model, the parameters of each user channel are directly estimated from a sequence of packet transmission which can be mapped into a hidden Markov Model (HMM).
- We investigate the problem of NSP profit maximization that enables a flexible tradeoff between the revenue that depends on user QoS and the associated bandwidth cost. The user QoS takes into consideration of the true payload transmission instead of the PHY layer throughput, which is greatly affected by the ROHC parameter optimization per MBMS user group. We solve the problem of profit maximization via joint group assignment and ROHC parameter optimization with an efficient dynamic programming (DP) algorithm.

The rest of this paper is structured as follows. Section II introduces the preliminary background and the ROHC system framework. In Section III, we develop a Markovian model for each wireless link equipped with ROHC under both channel models, which lays the foundation of the ROHC parameter optimization. In section IV, we depict the NSP profit function which reflects both the true QoS based on payload delivery experienced by all our users as well as the bandwidth cost of MBMS groups. We formulate the profit maximization problem. To solve this problem efficiently, we develop a DP-based algorithm in Section V. Numerical results are provided in Section VI to demonstrate the performance gain

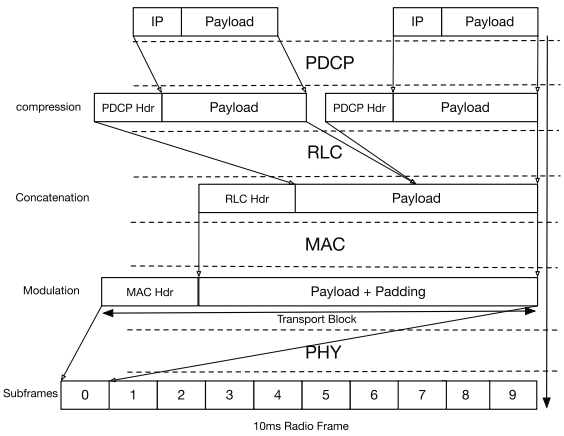


Fig. 1. LTE Protocol sublayer with IP packet transmission.

of our proposed MBMS scheme over conventional heuristic designs.

II. ROHC System Model

Let us first introduce the basic functions abilities of ROHC and its key elements. Fig. 1 illustrates the protocol stack of 3GPP LTE cellular services [21]. On top of Layer 2 of the LTE protocol stack, the PDCP sublayer is responsible for providing protocol interface between the cellular networks and the infrastructural internet. Within the PDCP sublayer, ROHC is responsible for converting the lengthy IP packet headers into the efficient cellular PDCP headers.

The ROHC module typically has three modes of operation: Unidirectional mode (U-mode), Bi-directional Optimistic mode (O-mode) and Bi-directional Reliable mode (R-mode) [12]. The U-mode is always the initial mode for all ROHC systems. ROHC can further transition into other two advance modes if configured and necessary. The major difference among the three modes lies in how ROHC compressor depends on the feedbacks from the decompressor to transitions between its three different states corresponding to three different types of headers. In U-mode, there is no feedback channel and the compressor depends on its internal time-out mechanism to change its state. All time-out parameters are determined by the NSP for the specific network in use. For both O-mode and R-mode, the compressor state transitions are controlled by feedbacks. As mentioned in [3], no dedicated feedback channel is used for ROHC in MBMS. Therefore, only the U-mode of ROHC is executed. According to [22], U-mode is less robust than R-mode and O-mode. However, it has higher efficiency in terms of channel resource usage and also does not suffer from feedback jitters. In this work, we study the open problem to optimize ROHC U-mode parameters for MBMS.

In ROHC, one of the most important objectives is header compression efficiency. The header fields in an IP packet can be generally classified into the static and the dynamic parts. The static parts (e.g. source and destination addresses) remain unchanged throughout the packet session and only needs to be successfully transmitted once for the decompressor to establish the context synchronization, while the dynamic

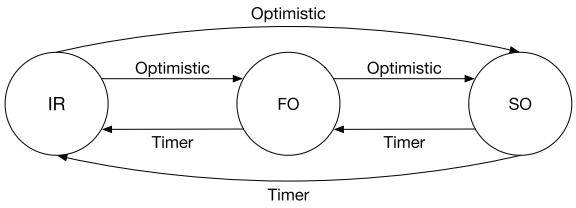


Fig. 2. Three states ROHC U-mode compressor. The upward state transitions are based on optimistic setup of NSP such as after transmitting L packets, the compressor goes to next state. The downward state transitions are based the timeout timer.

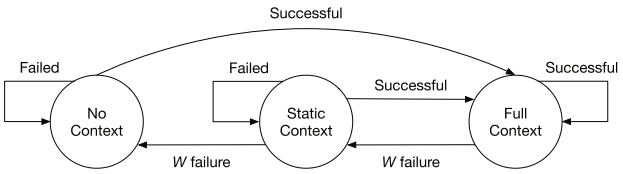


Fig. 3. Our three states ROHC decompressor model. The upward state transitions are based on the successful decompression. The downward state transitions are based on the repeat failed decompression. To stay at Full Context state, both successful decompression and less than W failed decompression are satisfied.

parts (e.g. IP serial number) changes with correlations between consecutive packets and therefore, can be compressed once the decompressor and the compressor establishes context synchronization. The ROHC compressor in the transmitter determines whether to compress the static/dynamic parts of a header and can be modeled into a finite-state machine (FSM) [12] with three states (Figure 2), each corresponding to a header type of different size.

The ROHC compressor always starts in the Initialization and Refresh (IR) state, where an uncompressed header is transmitted and both the static and dynamic context can be (re)established if the transmission is successful. In the First Order (FO) state, the compressor compresses the static field of the header such that only the dynamic context can be (re)established in a successful transmission. In the Second Order (SO) state, both the static and the dynamic fields are fully compressed [12].

Correspondingly, the ROHC [12] decompressor can also be described as a FSM with three states (Figure 3). The decompressor always starts with the No Context (NC) state, in which it has established the context synchronization for neither the static nor dynamic fields and can only decompress IR packets. After the first IR packet from the compressor is decompressed successfully, the decompressor establishes the context synchronization and enters the Full Context (FC) state where it can decompress IR, FO and SO headers. In case of consecutive packet losses due to the poor channel condition, the decompressor will lose the context synchronization with the compressor and can no longer decompress SO headers, thereby enters the Static Context (SC) state. From the SC state only a successful decompression of an IR or a FO header can re-establish full context synchronization for the decompressor. In case of further decompression failures, the decompressor will eventually transition downwards to the NC state.

An important header compression algorithm is Windowed Least Significant Bits (W-LSB) algorithm [23], which is so

designed in order to achieve a balance between compression efficiency and robustness against packet losses. Different from the conventional LSB, the compressor maintains a window of W reference values and uses the least significant bits enough to uniquely determine the field with any of the reference values. As a result, W-LSB can tolerant up to W consecutive transmission failures over the wireless channel. However, once the number of successive transmission failures exceeds W, the decompressor will not be able to decompress the W-LSB encoded and thus entering the SC state, from which only the successful transmission of an IR or FO packet can re-establish full context synchronization and drive the decompressor back to the FC state.

In practice, the ROHC compressor needs to alternately transmit the three types of headers in order to achieve a high compression efficiency without introducing many extra packet losses due to the context Out-of-Synchronization (OoS) caused by insufficient context refreshment. In general, if the compressor is confident that the decompressor has established full context, it will transmit the shortest SO headers. Conversely, the compressor will transmit a few of the longer IR/FO packets until it is confident enough that the decompressor has (re)established context synchronization [12]. As mentioned previously, in MBMS only U-mode ROHC is allowed since it is difficult to coordinate feedbacks from multiple users [24]. The conventional approach to control the state transition of the U-mode ROHC compressor is to use a timeout mechanism, i.e. the compressor transmits a fixed number of packets with a certain type of headers and then transit to another state [12]. If the user channel is generally good, optimizing timeout parameters enable the compressor to remain in the SO state for substantial portion of time, thereby achieving high packet transmission efficiency in terms of both channel bandwidth and transmission power, as discussed before.

We will investigate the timer parameter optimization of the U-mode ROHC compressor, which has been shown by the experimental results in [25]–[30] to be able to greatly affect the efficiency. Consider the long-term behavior of an ROHC compressor, the timer parameters are defined as the average duration of consecutive transmission of IR, FO and SO packets, denoted as D_{ir}, D_{fo}, and D_{so}, respectively. Note that in MBMS applications, the parameter optimization must take into account of the diversity of the users' channels, which we will introduce in the next section.

III. Markovian System Model for an ROHC Link

In this section, we develop the analytical framework for an ROHC link composed of a packet channel and a compressor/decompressor pair.

A. ROHC Channel Model

At the PDCP sub-layer, the channel model should capture the protocol level connection linking the compressor and decompressor. In this work, we adopt two channel models that facilitate analysis on the performance of ROHC. The first channel model is a more idealized channel model with a strong assumption that is easy to deal with. We then investigate a more practical channel model to establish a more comprehensive ROHC system model to provision optimized MBMS service groups.

1) PM-Model: PDCP Channel Derived From PHY Channel [20]: Our first modeling approach derives the PDCP-layer packet channel from the statistical Channel State Information (CSI) of the PHY fading channel [31]. Consider the flat fading PHY channel between the base station (BS) and n-th user, $n = 1, \ldots, N$

$$y = \alpha_n h_n x + \upsilon, \tag{1}$$

in which x is the transmission signal, υ is the channel noise, and the channel consists of a large scale fading factor α_n and a small-scale fading factor h_n with the following assumptions:

- For each user n, α_n is assumed to be stationary throughout the entire ROHC session, whereas h_n experiences block fading across different ROHC packets. In our simulation, we assume α_n follows i.i.d log-normal distribution, whereas h_n follows standard complex normal distribution (Rayleigh fading) as in [31]. Nevertheless, our analytical framework is compatible with other fading distribution assumptions.
- All users experience i.i.d. additive white Gaussian noises (AWGN), i.e. $\upsilon_n \sim \mathcal{CN}(0, \sigma^2)$.

Based on this PHY fading channel, the ROHC channel model can be modeled as a finite state Markov chain with K states [32], which is a generalization from the rather popular 2-state Gilbert-Elliot (GE) channel model [33]. Assuming that all the users share the same maximum Doppler frequency f_m and the same packet period T_p and selecting the same K to characterize the ROHC channels for all the users, a K-state Markov channel model for user n can be derived as in [31]. The resulting Markov packet channel model can be fully characterized with

- A K-by-K state transition matrix $\mathbf{T}_{ch}$, where $P_{ij} \triangleq \{\mathbf{T}_{ch}\}_{ij}$ represents the state transition probability from channel state i to state j as defined by [31, eqs. (10) and (11)]. Note that $\mathbf{T}_{ch}$ is the same for all the n users since it is solely determined by f_m and T_p.
- A K-by-1 vector $\boldsymbol{\rho}_n$, where $1 - \rho_{nk}$ represents the packet error rate (PER) for user n under channel state k, $k = 1, \ldots, K$. This can be derived from the symbol error rate defined in [31, eq. (12)], which is then converted to bit error rate (BER) for a given modulation scheme, before mapping to the corresponding packet error rate following [34].

Note that $\mathbf{T}_{ch}$ is fixed given K, f_m and T_p, while $\boldsymbol{\rho}_n$ is a function of α_n. We name this PDCP model as the PM (PHY-mapping) channel model.

2) FK-Model: PDCP Channel Model Estimated From Feedbacks: The other modeling approach is to assume that the PDCP channel follows a Markov process whose parameters can be directly estimated from a sequence of Automatic Repeat-reQuest (ARQ) feedbacks. This channel model is named as FK-model. The motivation of this approach is that, in a practical LTE mobile network, the PDCP sublayer may not know the fading characteristics of the PHY layer and the error correction schemes in the MAC/PHY layer, or it is simply too complicated to derive the PDCP layer channel from the PHY

channel model considering the functionalities of the RLC layer and MAC layer in between. With the FK-model, the ROHC entity is able to estimate the PDCP packet channel directly during the negotiation stage.

Similar to the PM-model, we assume that the PDCP channel is a finite state Markov chain with $K-$states, characterized by the state transition matrix $\mathbf{T}_{ch}$ and the transmission probabilities ρ_k, $k = 1, \ldots, K$. Since the ARQ feedbacks can only indicate whether a packet transmission is successful or not, the PDCP channel appears to be a Hidden Markov Model (HMM) to the ROHC entity. Therefore, the estimation of $\mathbf{T}_{ch}$ and $\boldsymbol{\rho}_n$ from a sequence of ARQ feedbacks can be implemented with the classic HMM learning techniques such as Baum-Welch algorithm [35]. A more practical approach, however, is to define a finite number of M Markov channel models, i.e.

$$\mathcal{M} = \{(T_{ch_1}, \rho_1), (T_{ch_2}, \rho_2), \ldots, (T_{ch_M}, \rho_M)\} \tag{2}$$

and perform a codebook selection based on the observation sequence $\boldsymbol{o}$ with a maximum likelihood criterion.

B. ROHC System Model

Now we are in a good place to develop the system model for a ROHC link, which consists of a compressor, a packet channel and a decompressor. For the compressor, as discussed in [12], the impact of FO state is rather limited. Thus, we consider a simplified compressor transmitting only IR and SO headers. Consequently, our timer-based compressor can be characterized with two parameters D_{ir} and D_{so}. In order to facilitate theoretical analysis and optimization, we approximate the timer-controlled compressor with a two-state Markov chain whose transition probabilities are optimized. The transition probabilities of the surrogate Markov compressor are denoted as P_{is} and P_{si}, which represent the transition probabilities from state IR to state SO and from state SO to state IR, respectively. In order to develop a one-to-one mapping between the Markov compressor and the timer based compressor, we match the expected duration of consecutive IR and SO states for the Markov compressor to D_{ir} and D_{so}, which leads to

$$D_{ir} = P_{is}^{-1}, D_{so} = P_{si}^{-1} \tag{3}$$

Therefore, the Markov compressor can be defined by the transition matrix:

$$\mathbf{T}_c = \begin{bmatrix} 1 - P_{is} & P_{is} \\ P_{si} & 1 - P_{si} \end{bmatrix} \tag{4}$$

With this approximation, we imitate the deterministic/periodic behavior of the timer-based compressor with the stochastic/stationary behavior of the Markov compressor, which greatly reduces the size of the state space and therefore facilitates a simpler analytical framework. In the following we will discuss about the optimization of P_{is} and P_{si}, which will then be mapped to the optimal D_{ir} and D_{so}.

Corresponding to our simplified compressor model, the decompressor is also simplified to have only the NC and FC states [33]. Since the compression of the dynamic fields relies on W-LSB which can tolerate up to W consecutive

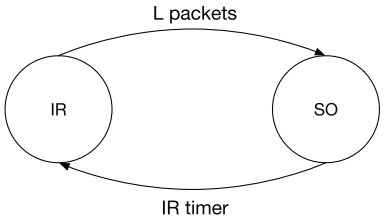


Fig. 4. Simplified ROHC System Model of Compressor only with IR and SO states.

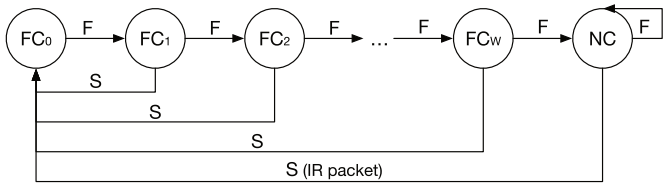


Fig. 5. Markov chain model for the ROHC U-mode decompressor with the window size W. The 'F' means transmission is failed and the 'S' means transmission is successful.

packets loss before losing context synchronization, the FSM model of the decompressor is fully depicted in Fig. 5. In this state-transition diagram, FC_w denotes that the decompressor is still in FC state but the last w packets have been lost. The symbol 'S' and 'F' denote the successful and failed packet transmission, respectively. From any of FC_w state, $w \leq W$, once a packet with either an IR or a SO header is transmitted successfully, the decompressor would be able to update its context and return to the FC_0 state. However, as explained in Section II, if more than W packets are lost in a roll, the decompressor will lose context and enters the NC state, from which only the successful transmission of an IR packet will enable the decompressor to re-establish context synchronization and enters the FC_0 state. We denote the set of states for the decompressor as

$$\mathcal{S}_d = \{FC_0, FC_1, \ldots, FC_W, NC\} \tag{5}$$

To summarize, the ROHC link for the n-th user can be modeled as a Markov chain which is composed of the Markov channel $\mathbf{T}_{ch}$, the Markov compressor $\mathbf{T}_c$, and the $(W+2)$-state FSM model for the decompressor. This aggregated Markov chain is defined with a $2(W+2)K$-by-$2(W+2)K$ state [19] transition matrix:

$$\mathbf{T}_{\text{sys}} = \begin{bmatrix} (1 - P_{is})\mathbf{T}_{\text{ir}} & P_{is}\mathbf{T}_{\text{ir}} \\ P_{si}\mathbf{T}_{\text{so}} & (1 - P_{si})\mathbf{T}_{\text{so}} \end{bmatrix} \tag{6}$$

in which

$$\mathbf{T}_{\text{ir}} = \begin{bmatrix} \mathbf{T}_{s,n} & \mathbf{T}_{f,n} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{T}_{s,n} & \mathbf{0} & \mathbf{T}_{f,n} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{s,n} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{f,n} \\ \mathbf{T}_{s,n} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{f,n} \end{bmatrix} \tag{7}$$

$$\mathbf{T}_{\text{so}} = \begin{bmatrix} \mathbf{T}_{s,n} & \mathbf{T}_{f,n} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{T}_{s,n} & \mathbf{0} & \mathbf{T}_{f,n} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{s,n} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{f,n} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{ch} \end{bmatrix} \tag{8}$$

Note that $\mathbf{T}_{s,n} = \mathbf{T}_{ch}\text{diag}(\boldsymbol{\rho}_n)$ and $\mathbf{T}_{f,n} = \mathbf{T}_{ch} - \mathbf{T}_{s,n}$. $\mathbf{T}_{\text{ir}}$ and $\mathbf{T}_{\text{so}}$ denotes the transition matrix of the decompressor and the channel when an IR packet or a SO packet is transmitted, respectively.

We define a ROHC performance measurement for each user which can be evaluated with the stationary distribution of the Markov model of the ROHC system. It is the transmission efficiency which is defined as the ratio between the successful decompressed payload bytes and the total transmitted bytes in the long term:

$$\eta = \frac{L_{\text{p}}\boldsymbol{\pi}_{\text{FC}_0}}{L_{\text{IR}}\boldsymbol{\pi}_{\text{IR}} + L_{\text{SO}}\boldsymbol{\pi}_{\text{SO}}} \tag{9}$$

where L_{p}, L_{IR} and L_{SO} represent the payload length per packet, the full packet length with an IR header, and the full packet length with a SO header, respectively. $\boldsymbol{\pi}_{\text{FC}_0}$ denotes the marginal stationary probability for the decompressor to be in the FC_0 state, whereas $\boldsymbol{\pi}_{\text{IR}}$ and $\boldsymbol{\pi}_{\text{SO}}$ represent the stationary probabilities for the Markov compressor to be in IR and SO states, respectively. To maximize the efficiency η, we need to evaluate the optimal ROHC compressor parameters P_{is} and P_{si} under specific channel condition determined by $\mathbf{T}_{ch}$ and $\boldsymbol{\rho}$. Recall that in our PM-model $\boldsymbol{\rho}$ depends on the large scale fading factor α_n and is hence different for all the users. In the FK model, both $\mathbf{T}_{ch}$ and $\boldsymbol{\rho}_n$ can be estimated from a sequence of ARQ feedbacks during the ROHC negotiation process. The objective function of our joint grouping and ROHC parameter optimization problem to be discussed in the next section will be closely related to η.

In the next section, we will consider a MBMS network service entity which consists of multiple ROHC links clustered into several multicast groups and formulate the joint grouping and ROHC optimization problem.

IV. The Joint MBMS Grouping and ROHC Optimization Scheme

As mentioned in Section I, the main design philosophy of our joint MBMS grouping and ROHC optimization scheme is to allow users with similar channel qualities to be clustered into the one multicast group, and then discriminatively select the parameters for the ROHC entity within each group in order to maximize the NPS profit. This problem is formulated as follows for the PM-channel model and the FK-channel model, respectively.

A. PM Channel Model

First, we consider the PM channel model in which the PDCP channel is derived from the PHY fading channel. The NSP divides the N users into several P-t-M MBMS groups served by individual P-t-M bearers. We assume that:

- The NSP allocates the same amount of bandwidth resources for each MBMS group at the cost of a. Without loss of generality, if users are divided into G groups, the total cost is $w(G) = aG$. Note that our solution approach is compatible with any choice of $w(G)$ which can be reasonably assumed to be a monotonically increasing function of G.
- The NSP collects revenue from each user within a P-t-M group based on the delivered packet QoS, defined as the payload rate r_g which can be adjusted by the MBMS application for each individual P-t-M group. The payload rate r_g must be supported by the channels between the BS and all users in group g. The revenue collected from each user in group g is defined as $f(r_g)$, where $f(\cdot)$ is a monotonically increasing function. Without loss of generality, we consider the case of a linear function $f(r_g) = cr_g$, where c is the positive pricing factor.
- The statistical CSI of all the N users are known to the NSP.

Specifically, denote $\mathcal{U} = \{1, \ldots, N\}$ as the set of N users and a G-partition of $\mathcal{U}$ as $\mathcal{C}_G = \{\mathcal{G}_1, \ldots, \mathcal{G}_G\}$ where $\mathcal{G}_g$ represents group g. For user n, the ergodic capacity supported by the channel is simply

$$r_n = \mathbb{E}_{h_n}\left[\log(1 + \|\alpha_n\|^2\|h_n\|^2)\right] \triangleq r(\alpha_n) \tag{10}$$

which is based on the Gaussian input assumption with $\sigma^2 = 1$. Also, a generalization to the more practical assumption of finite-state alphabet input signal is readily available [36]. The ROHC compressor for group g is modeled as a Markov chain as explained in Section III.B, which is characterized by a pair of parameters $\mathbf{p}_g = [p_{is}, p_{si}]$. Although all the users in one group share the ROHC compressor, the different channel qualities represented by α_n will lead to different ergodic capacity and transmission efficiency experienced by different users. The payload rate for MBMS group g must be adjusted so that it is supported by all the users in the group, i.e.

$$r_g \le r_n\eta(\alpha_n, \mathbf{p}_g), \quad \forall n \in \mathcal{G}_g \tag{11}$$

where $\eta(\alpha_n, \mathbf{p}_g)$ denotes the PDCP transmission efficiency function defined in Eq. (9) as a function of α_n and $\mathbf{p}_g$.

Our design objective is to maximize the benefit of the NSP, which equals to the difference between the total revenue and total bandwidth cost, by choosing the optimal group number G, user assignment $\mathcal{C}_G$ and the ROHC parameter $\mathbf{p}_g$ for each group:

$$\begin{aligned}\max_{G,\mathcal{C}_G,\{\mathbf{p}_g\}_{g=1}^{G}} \quad & \sum_{g=1}^{G} f(r_g)|\mathcal{G}_g| - w(G) \\ \text{s.t.} \quad & r_g \le r_n\eta(\alpha_n, \mathbf{p}_g), \quad \forall n \in \mathcal{G}_g,\ g = 1, \ldots, G\end{aligned} \tag{12}$$

B. FK-Model

For the FK-model of PDCP channel, the estimate of the model parameters based on PDCP feedbacks will be available for the NSP. Once the parameters $\mathbf{T}_{ch}$ and $\boldsymbol{\rho}$ are estimated for each user and the QoS requirement r_n is given, the joint grouping and ROHC optimization problem formulation remains nearly identical to that of the PM-model. In the next section, we will present an efficient DP-based solution for Eq. (12) which applies to both PM-model and FK-model.

V. A Dynamic Programming Solution

In this section, we propose an efficient algorithm, based on DP, to solve Eq. (12) with polynomial complexity.

For now assume that the total number of MBMS radio bearers (or groups) G is fixed. Note that both r_n and $\eta(\alpha_n, \mathbf{p}_g)$ given a fixed $\mathbf{p}_g$ are monotonically increasing with respect to α_n. Consequently, Eq. (12) with fixed G can be re-written into the following un-constrained optimization problem

$$\max_{\mathcal{C}_G} \sum_{g=1}^{G} \left[\max_{\mathbf{p}_{m_g}} f(r_{m_g}, \eta(\alpha_{m_g}, \mathbf{p}_{m_g})) |\mathcal{G}_g| \right] \tag{13}$$

where $m_g = \arg\min_{n \in \mathcal{G}_g} \alpha_n$ represents the user with the poorest channel condition in group g. Eq. (13) is based on the fact that the total revenue collected from a group depends on the user with the worst channel (i.e. the smallest α), and the ROHC parameter for the group should be optimized with respect to this user since all other users will always support a larger payload rate. Denote $\hat{\mathbf{p}}_n = \arg\max_{\mathbf{p}_n} \eta(\alpha_n, \mathbf{p}_n)$, i.e. the ROHC parameter optimized for user n. Since the revenue function $f(\cdot)$ is monotonically increasing w.r.t. the payload rate which is in turn monotonically increasing w.r.t the transmission efficiency η given r_n as in [37], we define

$$\begin{aligned} q_n &\triangleq \max_{\mathbf{p}_n} \; f(r_n, \eta(\alpha_n, \mathbf{p}_n)) \\ &= f(r_n, \eta(\alpha_n, \hat{\mathbf{p}}_n)) \end{aligned} \tag{14}$$

with which Eq. (13) can be rewritten as

$$\max_{\mathcal{C}_G} \sum_{g=1}^{G} q_{m_g} |\mathcal{G}_g| \tag{15}$$

Eq. (15) suggests that the optimization of ROHC parameters and user grouping for different P-t-M bearers can be decoupled by first optimizing the ROHC parameters for each user individually, followed by an optimized grouping scheme. To solve Eq. (15), we first present the following lemma which states that an optimal grouping scheme can be achieved by ordering and thresholding $\{\alpha_n\}$:

Lemma 1: Consider a user grouping denoted by $\mathcal{C}_G = \{\mathcal{G}_1, \ldots, \mathcal{G}_G\}$, in which there exists a $\mathcal{G}_i$ with $|\mathcal{G}_i| \geq 2$. If there is a user $v \in \mathcal{G}_j$, such that $\alpha_u > \alpha_v$ for some $u \in \mathcal{G}_i$, $i \neq j$ and $\alpha_{m_i} < \alpha_{m_j}$, then exchanging u and v between groups i and j results in a larger or equal total profit.

proof: As $\alpha_u > \alpha_v \geq \alpha_{m_i} < \alpha_{m_j}$,

- If $m_j \neq v$, after the exchange we have $m'_i = m_i$, $m'_j = m_j$, the total profit will not change.
- If $m_j = v$, after the exchange we have $m'_i = m_i$, $m'_j = u$. Since $\alpha_{m'_j} = \alpha_u > \alpha_v = \alpha_{m_j}$, the revenue collected from group j increases and the revenue collected from other groups would remain the same. Consequently, the total profit would increase from the exchange.

Without loss of generality, we assume that the users have been ordered as $\alpha_1 \leq \cdots \leq \alpha_N$. Lemma 1 suggests that an optimal solution to Eq. (15) can now be found by solving

$$\max_{1 < m_2 < \ldots < m_G \leq N} \sum_{g=1}^{G} q_{m_g}(m_{g+1} - m_g) - w(G) \tag{16}$$

where $m_1 = 1$, $m_{G+1} = N + 1$. Thus, the g-th group is identified as $\mathcal{G}_g = \{m_g, m_g + 1, \ldots, m_{g+1} - 1\}$, $g = 1, \ldots, G$. For a fixed number G, Eq. (16) amounts to $\max \sum_{g=1}^{G} q_{m_g}(m_{g+1} - m_g)$. Denote

$$R[n, g] = \max_{m_2, \ldots, m_g \in \{2, \ldots, n\}} \sum_{i=1}^{g} q_{m_i}(m_{i+1} - m_i) \tag{17}$$

i.e. the maximum revenue collected by grouping users $\{1, \ldots, n\}$ into g groups. Following Eq. (17), we can expand $R[n, g+1]$ as:

$$\begin{aligned} &R[n, g+1] \\ &\quad = \max_{m_2, \ldots, m_{g+1}} q_{m_1} n_1 + \cdots + q_{m_{g+1}} n_{g+1} \\ &\quad = \max_{m_{g+1}} \max_{m_2, \ldots, m_g | m_{g+1}} q_{m_1} n_1 + \cdots + q_{m_g} n_g + q_{g+1} n_{g+1} \end{aligned} \tag{18}$$

where $n_i = m_{i+1} - m_i$ is the number of users in group i and denote $m_{g+2} = n + 1$. Therefore, Eq. (18) can be re-defined recursively as:

$$R[n, g+1] = \max_{j: g \leq j \leq n-1} R[j, g] + q_{j+1}(n - j) \tag{19}$$

and the boundary conditions are:

$$R[g, g] = \sum_{i=1}^{g} q_i; \;\; R[n, 1] = n q_1 \tag{20}$$

DP is used to solve our problem since it computes the exact solution efficiently according to Lemma 1. It is noted that $R[N, G]$ is exactly the maximum revenue that can be collected by grouping the N users into G groups. Consequently, Eq. (19) and Eq. (20) together define a DP solution to Eq. (13). At the same time, the indices $\{m_g\}|_{g=1}^{G}$ representing an optimal grouping scheme can be re-constructed via backtracking [38, Ch.15]. This DP solution can be readily modified for Eq. (12), where $G = 1, \ldots, N$. In summary, our exact solution to the joint grouping and ROHC parameter optimization problem in Eq. (12) is summarized in Algorithm 1, whose time complexity is $O(N^3)$ and memory complexity is $O(N^2)$. In comparison, to find the same exact solution, an exhaustive search would have to go through of all possible grouping schemes, which amounts to Bell number [39]. As an example, we illustrate the time complexity tendency between DP-based algorithm and exhaustive search in Fig. 6.

VI. Simulation and Numerical Results

A. Test Setup

In this section, we present numerical results to evaluate the performance of our joint MBMS grouping and ROHC parameter optimization scheme for both the PM and FK channel models.

Our joint MBMS grouping and ROHC parameter optimization scheme is compared with the two conventional MBMS schemes, namely using N individual unicast bearers (P-t-P) and a single multicast bearer (P-t-M) to serve the N MBMS users. We also compare our proposed MBMS scheme with two heuristics MBMS grouping schemes, namely K-means and uniform clustering. In first heuristic scheme, the N users are clustered into K groups by applying the conventional

Algorithm 1 Maximum-Profit Joint MBMS Grouping and ROHC Parameter Optimization

1: Order the N users such that $\alpha_1 \leq \ldots \leq \alpha_N$.
2: Optimize the ROHC parameter for each user with

$$\hat{\mathbf{p}}_n = \arg\max_{\mathbf{p}_n} \eta(\alpha_n, \mathbf{p}_n)$$

for $n = 1, \ldots, N$, and evaluate the corresponding maximum revenue with Eq. (14).
3: Initialize $R[n, g]$ as per Eq. (20) and $m[n, g]$ as

$$m[g, g] = g,\ m[n, 1] = 1$$

for $n = 1, \ldots, N,\ g = 1, \ldots, n$.
4: **for** $g = 2$ **to** N **do**
5: **for** $n = g + 1$ **to** N **do**
6: Evaluate

$$R[n, g+1] = \max_{j:g\leq j\leq n-1} R[j, g] + q_{j+1}(n-j)$$
$$m[n, g+1] = \arg\max_{j:g\leq j\leq n-1} R[j, g] + q_{j+1}(n-j)$$

7: **end for**
8: **end for**
9: $\hat{G} = \arg\max_G R[N, G] - w(G)$.
10: Set $m_{G+1} = N + 1$, $m_1 = 1$, $n = N$.
11: **for** $g = \hat{G}$ **down to** 1 **do**
12: $m_g = m[n, g]$, $n = m[n, g] - 1$.
13: **end for**
14: $\mathbf{p}_g = \hat{\mathbf{p}}_{m_g}$ and assign user $m_g, \ldots, m_{g+1} - 1$ to group g, for $g = 1, \ldots, \hat{G}$.

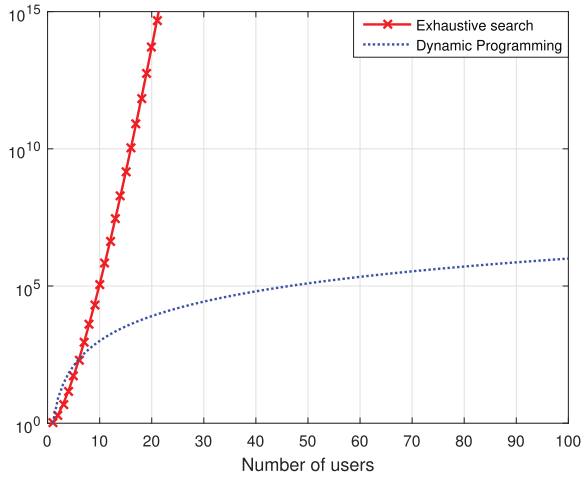


Fig. 6. The time complexity tendency between DP and exhaustive search.

K-means algorithm on $\alpha_n, n = 1, \ldots, N$. In the second heuristic scheme, each group has same number of users. The total profits of different grouping schemes are compared under different service cost a, number of users N, and channel distribution. We further evaluate the robustness of our proposed MBMS scheme by considering the effect of inaccurate statistical channel information.

Unless otherwise noted, all simulation tests are based on following parameter settings and evaluated over 500 Monte Carlo random runs.

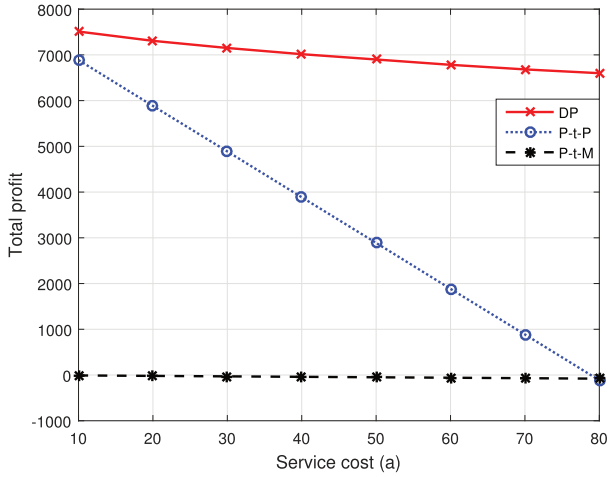


Fig. 7. Total profit versus different cost per ROHC bearer for $N = 100$ users achieved by the optimal DP scheme and pure unicast and one group multicast. To maintain a positive total profit of P-t-P transmission, the highest cost is set to $a = 80$ in this figure.

- $\{\alpha_n\}|_{n=1}^{N}$ are randomly generated from i.i.d. log-normal distribution of mean $m = 10$ and variance $v = 100$.
- The number of states for the Markov channel for all users is set to $K = 3$. Also, we set the maximum Doppler frequency and one packet time period as $f_m T_p = 0.0338$ [31].
- The maximum allowable number of consecutive header losses by WLSB is set to $W = 4$.
- The cost per MBMS bearer $a = 80$.
- The total number of users $N = 100$.
- The number of PDCP models in the channel codebook of the FK channel model $M = 5$.

B. NSP Profit From P-t-M Assignment

First, we compare the performance of our proposed joint MBMS grouping and ROHC parameter optimization scheme with that of the two conventional MBMS configurations, namely using one single P-t-M bearer or N P-t-P bearers. The performances are measured as the total profit defined in Eq. (12) versus varying a. As shown in the Fig.7, when the cost per bearer is low, P-t-P solution is nearly optimum as our DP solution. However, as the bandwidth cost grows with a, P-t-P unicast will be increasingly costly such that the NSP profit will decrease linearly. Although the profit from our proposed DP-based grouping scheme also tends to drop, the reduction is rather mild. When a pure P-t-M radio bearer is set to serve all N users in the conventional MBMS service, the service provider can hardly gain any profit since the user with the poorest channel greatly limits the revenue of the entire group.

To further justify the benefit of our DP-based grouping scheme, it is compared against a number of heuristic methods. Our heuristic MBMS grouping methods are based on $K-$means clustering. In our comparison, we use two versions of the $K-$means clustering differed by the way to determine

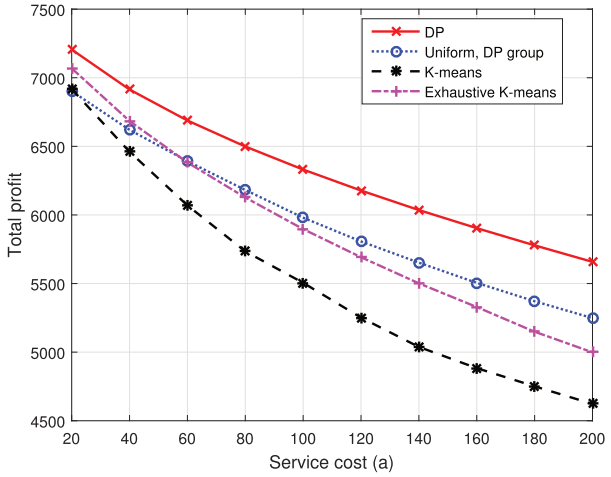


Fig. 8. Total profit versus different cost per ROHC bearer for $N = 100$ users achieved by the optimal DP scheme and the two heuristic schemes.

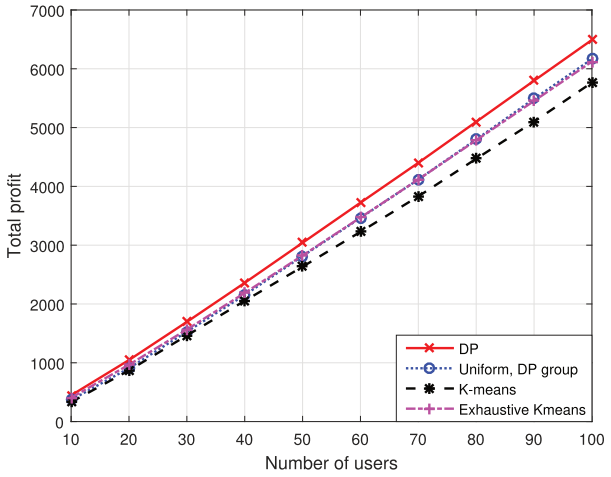


Fig. 9. Total profit versus different number of users and fixed number of groups $G = 4$.

TABLE I

GROUP SIZE FOR DIFFERENT COST a

a	20	40	60	80	100
DP	17.366	12.584	10.220	8.876	7.986
$K-$means	22.062	16.638	13.984	12.226	10.832
a	120	140	160	180	200
DP	7.330	6.840	6.472	6.106	5.892
$K-$means	9.992	9.254	8.750	8.192	7.640

the total number of groups G. In the first version, the number of clusters for the $K-$means method is chosen as the optimal G determined by our DP algorithm. In the second version of $K-$means, G is selected with an exhaustive search to maximize the resulting total profit. We also propose a simple heuristic scheme named uniform grouping, in which the users are ordered and grouped into G groups, each of which has the same number of users. Once again, we assume that G is determined by our DP-based solution.

C. PM-Model

First, we consider the PM-model. Fig. 8 presents the comparison of different grouping schemes in NSP total profit versus the cost per MBMS bearer a. When the cost a is low, the DP algorithm shall allow more P-t-M bearers to maximize total profit. As a increases, the number of groups for DP-based algorithm would naturally decrease and the profits of all schemes would drop as well. In these test results, the proposed DP algorithm always outperforms the three heuristic MBMS grouping schemes. Better insights can be obtained when we provide the average number of groups determined by both the DP algorithm and the exhaustive search $K-$means algorithm in Table I for each point in Fig. 8. For each value of a, the DP algorithm results in a higher profit by employing a lower number of groups. Thus our proposed method has a higher spectrum efficiency than the other heuristic MBMS clustering schemes.

From [3], it is clear that the MBMS can provide service to new users trying to join the MBMS that is already transmitting in. Hence, a more practical problem is to see how P-t-M performance would vary as the number of multimedia

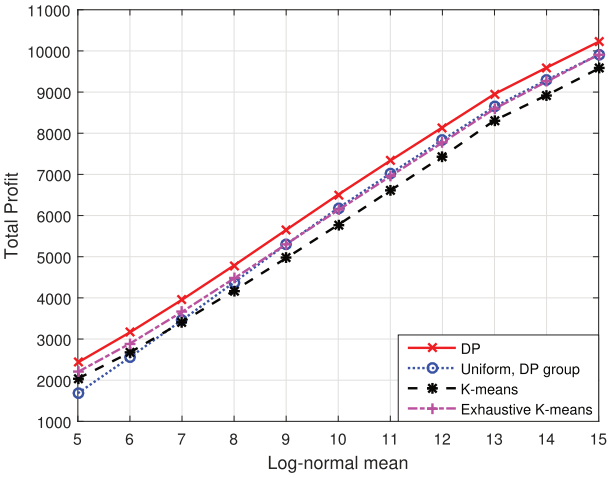


Fig. 10. Total profit versus different user channel condition for $N = 100$ users achieved by the optimal DP scheme and the two heuristic schemes.

subscribers group. In Fig. 9, we evaluate the performances in terms of total profit versus varying number of users for the same four MBMS grouping schemes. Naturally, the total profit of each scheme would increase with larger N. Nevertheless, the advantage of our proposed DP algorithm becomes more obvious as N grows.

In Fig. 10, the effect of different user channel condition are evaluated. The total profit is tested for log-normal CSI distributions with different parameters (mean and variance). Since the channel conditions improve with increasing mean, the overall profit grow for each P-t-M bearer assignment schemes as all users' channel condition become good and DP-based algorithm always provides a higher profit than the rest three heuristic schemes.

In both Fig. 9 and Fig. 10, the improvements between our DP-based algorithm and other methods are less than 10%. The reason is that the profit of our DP-based algorithm is evaluated by considering the user QoS which is ignored by other schemes such as exhaustive K-means. Although the gap between them is small, the heuristic methods can not ensure the user QoS and the DP-based algorithm still has a higher profit.

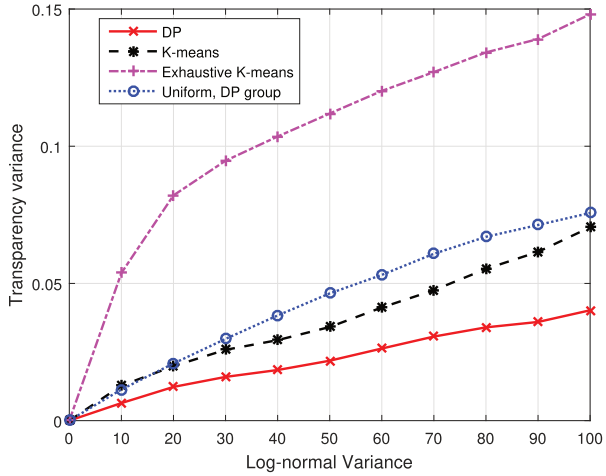


Fig. 11. The variance of within cluster packet loss rate caused by failed decompress versus different variance of log-normal distribution for $N = 100$ users. The mean is fixed $m = 10$.

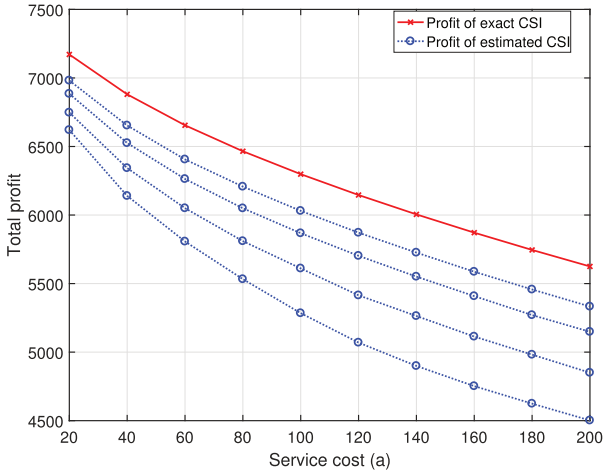


Fig. 12. Total profit versus different cost per ROHC bearer achieved with exact/estimated channel information for $\epsilon = 0.5$dB, 1dB, 1.5dB, 2dB from top to bottom.

D. Transparency

In this section, we evaluate another performance measurement named transparency [40], defined as the probability of successful decompression conditioned on a successful transmission, in order to gain some insight into the advantage of the DP-based grouping scheme. Transparency is an indicator of how well ROHC avoids introducing additional packet loss due to the compression which is represented as:

$$P_T = \frac{\boldsymbol{\pi}_{FC_0}}{\boldsymbol{\rho}^T \boldsymbol{\pi}_{ch}} \tag{21}$$

where $\boldsymbol{\pi}_{ch}$ is the stationary probability of the channel state transition matrix $\mathbf{T}_{ch}$. Fig. 11 illustrates the variance of the transparency within each MBMS group. From the figure, our proposed DP-based algorithm has a lower transparency variance which means the successful decompressed probability is higher for all users. This is because that DP-based algorithm always provides an optimal grouping scheme based on the user channel condition which can lead a lower out of synchronization probability between compressor and decompressor. In other heuristic methods, they may have similar profits at some points in the previous figures, however, they can not satisfy all users when the channel conditions are variable. The test shows that the DP-based algorithm always leads to the smallest variance of transparency for users within each P-t-M (MBMS bearer) or user cluster. Thus, for users assigned to the same MBMS bearer, our DP algorithm delivers the most consistent performance.

E. Channel Uncertainty

Next, we consider a practical issue involving the practical problem of inaccurate channel information available to the transmitter. In practice, the large scale fading factors α_n are not likely to be perfectly known by the BS of the NSP. Often, only inaccurate CSI are available to the BS. In this simulation, we model the channel uncertainty as:

$$\tilde{\alpha}_n\,(\text{dB}) = \alpha_n\,(\text{dB}) + \Delta\alpha_n\,(\text{dB}) \tag{22}$$

where $\Delta\alpha_n$ represents the channel estimation error. Assume that the estimation error is bounded by $|\Delta\alpha_n| \leq \epsilon$ as in [41]. We generalize our proposed joint grouping and optimization scheme in Eq. (12) into the following max-min criterion:

$$\max_{G, C_G, \{\mathbf{p}_g\}|_{g=1}^{G}} \sum_{g=1}^{G} f(r_g)|\mathcal{G}_g| - w(G)$$
$$\text{s.t.} \quad r_g \leq r_n \eta(\alpha_n - \epsilon, \mathbf{p}_g), \quad \forall n \in \mathcal{G}_g, \ g = 1, \ldots, G \tag{23}$$

which aims at maximizing the worst-case total profit. This robust version of the joint grouping and ROHC optimization can be solved by simply adopting our DP-based algorithm on $\{\tilde{\alpha}_n - \epsilon\}|_{n=1}^{N}$ instead of $\{\alpha_n\}|_{n=1}^{N}$. To evaluate its performance, we randomly generate $\Delta\alpha_n$ from a uniform distribution $U(-\epsilon, \epsilon)$. In Fig. 12, we compare the maximum total profit achieved by the DP-algorithm relying the true $\{\alpha_n\}|_{n=1}^{N}$, and the DP-algorithm relying on estimated $\{\tilde{\alpha}_n\}|_{n=1}^{N}$ based on the different values of ϵ. When the service cost is low, the DP-based algorithm can provide more profit for separating users to more groups. Therefore, there is no big gap between the estimated CSI and exact CSI at point $a = 20$ and $a = 40$. Also, the DP-based algorithm can adjust the grouping scheme to handle small error when the service cost is extremely large, such as $a = 200$. The simulation results suggest that our DP-based algorithm is quite robust against CSI uncertainty.

F. FK-Model

We next test our algorithm under the FK-models of PDCP channels. First, we evaluate the total profit according to variable cost a to test our proposed DP-based algorithm. Fig.13 shows the almost same result as the PM-model.

Considering the FK model, a specific problem is the effect of the number of Markovian channel template in the channel codebook on the MBMS performance, which is shown in Fig. 14. In our simulation, basic sample templates are uniformly selected based on the user distribution from the NSP. It seems that 4 or 5 PDCP templates are sufficient to

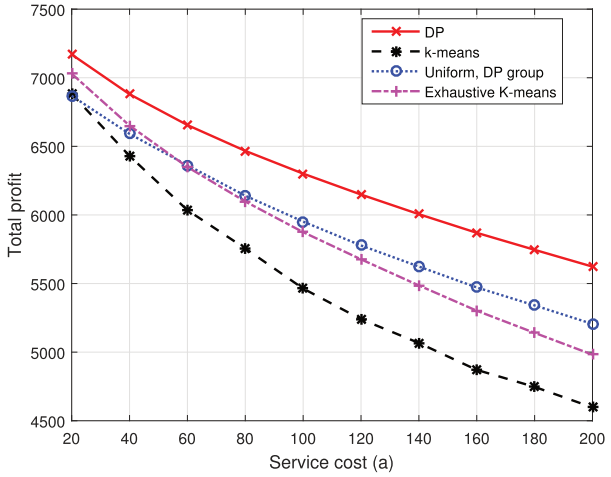


Fig. 13. Total profit versus different cost per ROHC bearer for $N = 100$ users achieved by the optimal DP scheme, $K-$means and the two heuristic schemes.

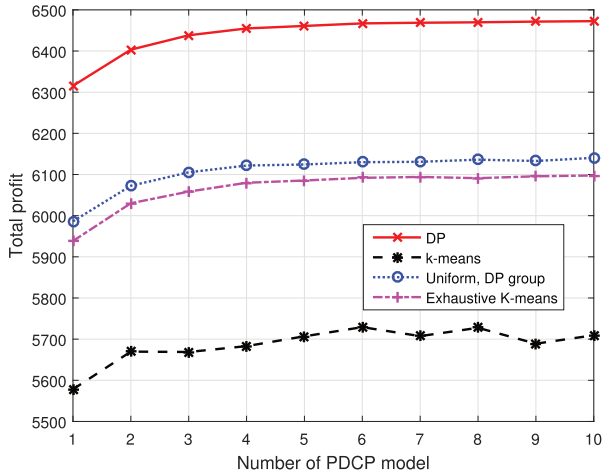


Fig. 14. Total profit versus different number of PDCP models for $N = 100$ users achieved by the optimal DP scheme, $K-$means, exhaustive $K-$means and uniform.

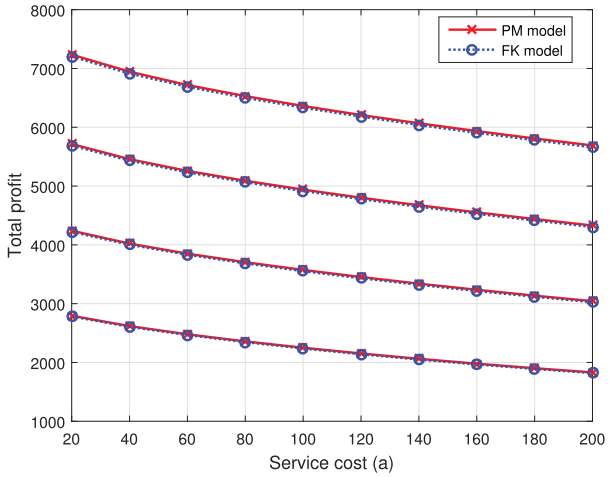


Fig. 15. The total profit versus different cost of two different channel model for $N = 40, 60, 80, 100$ users from the bottom up.

characterize the channels of 100 users in all four grouping schemes. Using more than 5 PDCP templates does not appear to improve the total profit.

Finally, we compare the performance of the two different channel models (PM-model vs. FK-model). As explained in Section III.A, the PM-model relies on the raw user CSI from the PHY layer, whereas the FK-model estimates its channel parameters by mapping the CSI from the PHY channels. To illustrate the performance gap of the practical FK-model versus the ideal -PM-model, we construct a PM-model with our default simulation settings, then randomly generate sequences of 1000 ARQ feedbacks, from which the PDCP-layer channel is estimated as in the FK-model. The exact and the estimated channel parameters are then used as input to Algorithm 1, respectively, and the grouping and ROHC parameter optimization results from two channel models are evaluated and compared with the same true channel model. The results are shown in Fig 15. From these Monte Carlo tests, we can see that the performance of the FK model does not degrade significantly with the in-exact estimated channel parameters as long as there are sufficient ARQ feedbacks.

VII. Conclusion

This work investigates the service provision of MBMS in 4G-LTE cellular networks to achieve high MBMS transmission rate and spectrum efficiency. We formulated a joint optimization problem of MBMS user assignment and ROHC parameters optimization to maximize the profit of network NSP under two different channel models. We developed a Markov model for each ROHC-enabled link in the MBMS network. Our design objective is to achieve the tradeoff between bandwidth usage cost and user QoS. We further derive an efficient dynamic programming algorithm to solve the optimization problem with demonstrated superior performance in simulations.

References

[1] *Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) Specification*, document 3GPP TS36.323, V13.2.1, Jul. 2016.

[2] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. Hoboken, NJ, USA: Wiley, 2005.

[3] *Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (RAN); Stage 2*, document 3GPP TS 25.346, V13.0.0, Jan. 2016.

[4] D. Gómez-Barquero, A. Fernandez-Aguilella, and N. Cardona, "Multicast delivery of file download services in evolved 3G mobile networks with HSDPA and MBMS," *IEEE Trans. Broadcast.*, vol. 55, no. 4, pp. 742–751, Dec. 2009.

[5] L. Militano, M. Condoluci, G. Araniti, and A. Iera, "Multicast service delivery solutions in LTE-advanced systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 5954–5958.

[6] H. Won *et al.*, "Multicast scheduling in cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4540–4549, Sep. 2009.

[7] A. Alexiou, C. Bouras, and E. Rekkas, "An improved MBMS power counting mechanism towards long term evolution," *Telecommun. Syst.*, vol. 43, nos. 1–2, pp. 109–119, 2010.

[8] I. G. Fraimis, V. D. Papoutsis, K. Ioannou, and S. A. Kotsopoulos, "Power efficient radio bearer selection for mbms users in soft handover status," in *Proc. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2009, pp. 2571–2575.

[9] W. Ji, P. Frossard, B.-W. Chen, and Y. Chen, "Profit optimization for wireless video broadcasting systems based on polymatroidal analysis," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2310–2327, Dec. 2015.

[10] J. F. Monserrat, J. Calabuig, A. Fernandez-Aguilella, and D. Gomez-Barquero, "Joint delivery of unicast and E-MBMS services in LTE networks," *IEEE Trans. Broadcast.*, vol. 58, no. 2, pp. 157–167, Jun. 2012.

[11] EFFNET AB, (Feb. 2004). "An introduction to IP header compression." Effnet AB. Mountain View, CA, USA. White Paper. [Online]. Available: http://http://www.effnet.com/

[12] C. Bormann *et al.*, *RObust Header Compression (ROHC): Framework and Four Profiles: RTP, UDP, ESP, and Uncompressed*, document RFC 3095, Jul. 2001.

[13] H. Woo, J. Kim, M. Lee, and J. Kwon, "Performance analysis of robust header compression over mobile WiMAX," in *Proc. IEEE 10th Int. Conf. Adv. Commun. Technol. (ICACT)*, vol. 3. Feb. 2008, pp. 1742–1746.

[14] M. Tomoskozi, P. Seeling, and F. H. Fitzek, "Performance evaluation and comparison of RObust Header Compression (ROHC) ROHCv1 and ROHCv2 for multimedia delivery," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 544–549.

[15] S. Traboulsi, W. Zhang, D. Szczesny, A. Showk, and A. Bilgic, "An energy-efficient hardware accelerator for robust header compression in LTE-advanced terminals," in *Proc. IEEE 22nd Int. Conf. Field Program. Logic Applicat. (FPL)*, Aug. 2012, pp. 691–694.

[16] D. E. Taylor, A. Herkersdorf, A. Doring, and G. Dittmann, "Robust header compression (ROHC) in next-generation network processors," *IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 755–768, Aug. 2005.

[17] S. Kalyanasundaram, V. Ramachandran, and L. M. Collins, "Performance analysis and optimization of the window-based least significant bits encoding technique of ROHC," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 4681–4686.

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

[19] C. Jiang, W. Wu, and Z. Ding, "IP packet header compression and user grouping for LTE multimedia broadcast multicast services," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Jan. 2017, pp. 52–57.

[20] C. Jiang, W. Wu, and Z. Ding, "LTE multimedia broadcast multicast service provisioning based on robust header compression," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[21] A. Larmo, M. Lindström, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 52–59, Apr. 2009.

[22] A. Couvreur, L.-M. Le Ny, A. Minaburo, G. Rubino, B. Sericola, and L. Toutain, "Performance analysis of a header compression protocol: The ROHC unidirectional mode," *Telecommun. Syst.*, vol. 31, no. 1, pp. 85–98, 2006.

[23] EFFNET AB, (Feb. 2004). "The concept of robust header compression, rohc." Effnet AB, White Paper. [Online]. Available: http://http://www.effnet.com/

[24] B. Quinn and K. Almeroth, *IP Multicast Applications: Challenges and Solutions*, document RFC 3170, Sep. 2001.

[25] E. Martinez, A. Minaburo, and L. Toutain, "Cross layer ROHC compression for multicast video streaming," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2008, pp. 2878–2882.

[26] A. Maeder and A. Felber, "Performance evaluation of ROHC reliable and optimistic mode for voice over LTE," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–5.

[27] B. Hung, D. Defrancesco, B.-N. Cheng, and P. Sukumar, "An evaluation of IP header compression on the GIG joint IP modem system," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Ocrt. 2014, pp. 1484–1490.

[28] W.-C. Ang, T.-C. Wan, K. Kataoka, and C.-H. Teh, "Performance evaluation of robust header compression (ROHC) over unidirectional links using DVB-S testbed," *KEIO SFC J*, vol. 8, no. 2, pp. 21–36, 2008.

[29] F. H. P. Fitzek, S. Rein, P. Seeling, and M. Reisslein, "RObust header compression (ROHC) performance for multimedia transmission over 3G/4G wireless networks," *Wireless Pers. Commun.*, vol. 32, no. 1, pp. 23–41, 2005.

[30] A. Minaburo, L. Nuaymi, K. D. Singh, and L. Toutain, "Configuration and analysis of robust header compression in UMTS," in *Proc. 14th IEEE Pers., Indoor Mobile Radio Commun. (PIMRC)*, vol. 3. Sep. 2003, pp. 2402–2406.

[31] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[32] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[33] R. Hermenier, F. Rossetto, and M. Berioli, "On the behavior of RObust header compression U-mode in channels with memory," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3722–3732, Aug. 2013.

[34] H. Bischl and E. Lutz, "Packet error rate in the non-interleaved Rayleigh channel," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1375–1382, Feb. 1995.

[35] A. Stolcke and S. Omohundro, "Hidden Markov model induction by Bayesian model merging," in *Proc. Adv. Neural Inf. Process. Syst.*, 1993, pp. 11–18.

[36] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for finite-alphabet inputs over MIMO fading channels with statistical CSI," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3134–3148, Jun. 2012.

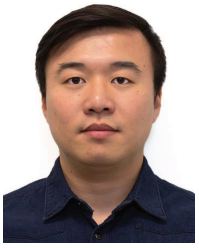
[37] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291–303, Feb. 2002.

[38] T. H. Cormen, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2009.

[39] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

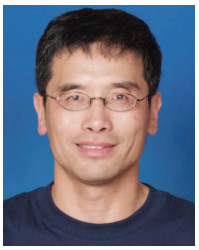
[40] C. Y. Cho, W. K. G. Seah, and Y. H. Chew, "A framework and source model for design and evaluation of robust header compression," *Comput. Netw.*, vol. 50, no. 15, pp. 2676–2712, Oct. 2006.

[41] A. Tajer, N. Prasad, and X. Wang, "Robust linear precoder design for multi-cell downlink transmission," *IEEE Trans. Signal Process*, vol. 59, no. 1, pp. 235–251, Jan. 2011.



Chen Jiang (S'08) received the B.S. degree in electrical engineering and automation from the Huazhong University of Science and Technology, Wuhan, China, in 2008, and the master's degree in electrical and computer engineering from Wichita State University, Wichita, KS, USA, in 2013. He is currently pursuing the Ph.D. degree in electrical and computer engineering at the Broadband Radio Access Technologies Laboratory, University of California at Davis, Davis, CA, USA. His research interests include wireless communication on broadcast/multicast and trans-layer designs.

Wenhao Wu (S'12) received the B.S. degree in electrical information science and technology from Tsinghua University, Beijing, China, in 2012, and the Ph.D. degree in electrical and computer engineering with the Broadband Radio Access Technologies Laboratory, University of California at Davis, Davis, CA, USA, in 2017. During his Ph.D. studies, he conducted cooperative research at the Missouri University of Science and Technology, Rolla, MO, USA. His research interests include communications and information theory with special emphasis on multiple-input multiple-output systems, signal processing for wireless communications, and trans-layer designs.



Zhi Ding (S'88–M'90–SM'95–F'03) received the Ph.D. degree in electrical engineering from Cornell University in 1990. From 1990 to 2000, he was a Faculty Member of Auburn University and the University of Iowa. He has held visiting positions at Australian National University, the Hong Kong University of Science and Technology, the NASA Lewis Research Center, and the USAF Wright Laboratory. He has active collaboration with researchers from several countries, including Australia, China, Japan, Canada, Taiwan, South Korea, Singapore, and Hong Kong. He is currently a Professor of engineering and entrepreneurship at the University of California at Davis, Davis, CA, USA.

Dr. Ding has been an active member of IEEE, serving on the technical programs of several workshops and conferences. He was a member of the Technical Committee on Statistical Signal and Array Processing and a member of Technical Committee on Signal Processing for Communications (1994–2003). He served as an IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS Steering Committee Member (2007–2009) and its Chair (2009–2010). He received the 2012 IEEE Wireless Communication Recognition Award from the IEEE Communications Society and is a co-author of the book *Modern Digital and Analog Communication Systems* (Fourth Edition, Oxford University Press, 2009). He is also an IEEE Distinguished Lecturer of the Circuits and Systems Society, 2004–2006, and the Communications Society, 2008–2009. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994–1997, 2001–2004, and the IEEE SIGNAL PROCESSING LETTERS 2002–2005. He was the Technical Program Chair of the 2006 IEEE Globecom.