

Delay-Throughput Enhancement in Wireless Networks with Multi-path Routing and Channel Coding

Keivan Ronasi, *Student Member, IEEE*, Amir-Hamed Mohsenian-Rad, *Member, IEEE*, Vincent W.S. Wong, *Senior Member, IEEE*, Sathish Gopalakrishnan, *Member, IEEE*, and Robert Schober, *Fellow, IEEE*

Abstract—Multi-path routing and adaptive channel coding are two well-known approaches that have been separately applied to wireless networks in order to improve the effective throughput. However, it is usually expected that achieving a high throughput would be at a noticeable cost of increasing the average end-to-end delay and causes major degradation in the overall network performance. In this paper, we show that a combination of multi-path routing and adaptive channel coding can improve throughput and reduce delay, and that it is possible to trade off delay for throughput and vice versa. This is in contrast to the general expectation that higher throughput can only be achieved with noticeable degradations in the end-to-end network delay. In this regard, we jointly formulate the end-to-end data rate allocation and adaptive channel coding (at the physical layer) within the general framework of network utility maximization (NUM). Depending on the choice of the objective function, we formulate two NUM problems: one aiming to maximize the aggregate network utility; another one aiming to maximize the minimum utility among the end-to-end flows in order to achieve fairness, which is of interest in certain vehicular network applications. Simulation results confirm that we can decrease the average delay significantly at the cost of a small decrease in throughput. This is achieved by maximizing the aggregate utility in the network when fairness is not the dominant concern. Furthermore, we also show that even when resource allocation is performed in order to provide fairness, we can still decrease the maximum end-to-end delay of the network at the cost of a slight decrease in the minimum throughput.

Keywords: Link reliability, multi-path routing, adaptive channel coding, delay-throughput trade-off, utility maximization, fairness, non-convex optimization, signomial programming.

I. INTRODUCTION

While most of the new multimedia applications have strict quality-of-service (QoS) requirements [1], the existing best-effort traffic delivery model cannot provide any service guarantee with respect to the minimum throughput and maximum delay of the end-to-end flows. Therefore, it is important to

design wireless networks with high performance in regard of delay and throughput.

Following the work by Kelly *et al.* [2], network utility maximization (NUM) has been widely used as a framework to systematically devise resource allocation strategies that can enhance the network performance subject to various capacity and QoS constraints [3]–[6].

It is known that multi-path routing improves the network performance by not only distributing the traffic over different links, but also by providing alternative paths for those sessions which are exposed to high bit error rates due to environmental conditions [7], [8]. The improvements lead to reducing network congestion, increasing throughput, and also higher energy efficiency [9]. On the other hand, adaptive channel coding (cf. [6], [10]) is used in the wireless networking context to improve the reliability of the transmissions, i.e., increasing the number of error-free delivered packets. Through adaptive channel coding, we provide higher resistance to errors in data packets by adding redundant bits. This in turn decreases the aggregate information sending rate on each link and correspondingly introduces a trade-off between throughput and reliability.

Recently, we investigated the trade-off between reliability and throughput in achieving the highest possible *effective throughput*, which is the end-to-end throughput that the receiver is able to receive [11]. We focused on *multi-path routing* wireless systems, where *adaptive channel coding* is also performed at the physical layer. However, we *did not* address the issue of delay in our earlier work. In this paper, we explicitly incorporate delay in the utility of each session and propose a joint data rate and coding rate allocation algorithm that leads to maximizing the network aggregate utility across all sessions. Our work complements the existing results in the literature as follows. The recent work by O'Neill *et al.* [12] used NUM with adaptive modulation to determine the optimal sending rates and transmit powers that maximize system performance. The trade-off between data rate, energy consumption, and delay is studied. However, O'Neill *et al.* did not incorporate delay into the utility function in their problem formulation [12] and the proposed design neither minimizes the delay nor provides a bound on end-to-end delay. On the other hand, Saad *et al.* [13] used the $M/G/1$ queueing model to estimate the delay as the summation of transmission delay and queueing delay. The same authors examined upper bounds on delay [14] but did not focus on

Manuscript was received on May 4th 2010, revised on Oct. 25th 2010, and accepted on Nov. 26th 2010. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The review of this paper was coordinated by Prof. Peter Langendoerfer.

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

K. Ronasi, V.W.S. Wong, S. Gopalakrishnan, and R. Schober are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada, email: {keivanr, vincentw, sathish, rschober}@ece.ubc.ca. A. H. Mohsenian-Rad is with the Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX, 79409, email: hamed.mohsenian-rad@ttu.edu.

delay reduction. In work by Kallitsis *et al.* [15], resources are allocated to maximize the throughput of the network and minimize the delay. Delay is modeled using network calculus and is incorporated directly into the utility function. Another research direction focuses on resource allocation to enhance the network performance by only minimizing the delay (e.g., Li *et al.* [16] and Kalyanasundaram *et al.* [17]). However, the impact of adaptive channel coding has not been considered in this context. On the other hand, channel coding is considered in [6]; but no analysis is performed related to delay. Finally, our problem is closely related to the recent work by Li *et al.* [18], which only addresses single-path routing within the context of wired networks or wireless networks with fixed capacity links. The contributions of this paper can be summarized as follows:

- We model a wireless network with several unicast data sessions, *multiple* routing paths for each session, and *adaptive channel coding* at the physical layer. To model the end-to-end delay, we use the average waiting time in an $M/D/1$ queueing system [19]. We then formulate the NUM problem of jointly finding optimal *sending rates* and *code rates* in the network to achieve the *maximum network utility* as a function of *throughput* and *delay*.
- We formulate two design optimization problems with and without fairness provisioning. In the former one, we aim to maximize the minimum utility in the network. In the latter case, we maximize the overall utility of the network. Fair resource allocation is of particular interest in vehicular networks in which moving vehicles frequently switch among stationary access points.
- To overcome the non-convexity due to channel coding, multi-path routing, delay and reliability consideration, we introduce new variables, constraints, and approximations in the original problem and reformulate it as a series of tractable *geometric programming* problems [20].
- We develop an iterative algorithm to solve the formulated problem. To the best of our knowledge, there has been no prior work on jointly improving throughput and delay in a *wireless multi-path* routing network with *channel coding* applied at the physical layer.
- Simulation results for random topologies show that, when fairness is not a concern, we can decrease the average delay by 60% at the cost of only a marginal ($< 0.1\%$) degradation in throughput. We also show that if fairness is addressed, we can decrease the maximum delay across the network by more than 35% with less than 12% decrease in minimum throughput.

Paper Organization: The system model and problem formulation are described in Section II. The delay-aware optimal data rate and coding rate allocation approach is introduced in Section III. The numerical results are shown in Section IV. The paper is concluded in Section V.

II. SYSTEM MODEL

A wireless network is modeled as a directed graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of nodes and $\mathcal{E}$ represents the set of wireless links, as it is shown in Fig. 1. For each unicast session

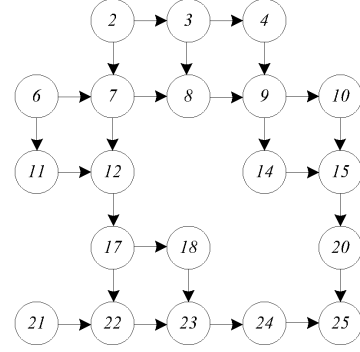


Fig. 1. A sample topology with five unicast multi-path data sessions. Data sessions are $(2 \rightarrow 15)$, $(7 \rightarrow 18)$, $(9 \rightarrow 20)$, $(12 \rightarrow 23)$, and $(3 \rightarrow 25)$. These sessions have 6, 1, 2, 2, and 4 available paths, respectively.

$i \in \mathcal{I}$, where $\mathcal{I} = \{1, 2, \dots, I\}$, the source and destination nodes are denoted by s_i and t_i , respectively. We define $\mathcal{K}_i$, with $K_i = |\mathcal{K}_i|$, as the set of all available routing paths from s_i to t_i . Moreover, for each session $i \in \mathcal{I}$, each $k = 1, \dots, K_i$, and each link $e \in \mathcal{E}$, we define

$$a_i^{ek} = \begin{cases} 1, & \text{if link } e \in k^{\text{th}} \text{ routing path for session } i, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For each session $i \in \mathcal{I}$, let α_i^k denote the data rate at source s_i on its k^{th} routing path, $k = 1, \dots, K_i$. Channel coding can improve the reliability over lossy wireless channels by adding redundant bits to data packets. For each link $e \in \mathcal{E}$, we define R_e as the link *coding rate*, i.e., the ratio of the number of data bits at the input of the encoder to the number of data plus redundant bits at the output. Notice that if channel coding is not performed on link e , then $R_e = 1$. Given the data rates at the sources $\alpha = (\alpha_i^k, i \in \mathcal{I}, k = 1, \dots, K_i)$ and the link coding rates $\mathbf{R} = (R_e, e \in \mathcal{E})$, the aggregate traffic load on each link $e \in \mathcal{E}$ is $u_e = \frac{1}{R_e} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k$. The smaller the coding rate R_e , the more redundant data is added to the transmitted packets on link $e \in \mathcal{E}$ leading to more reliable transmissions, i.e., transmissions with lower error probability. However, this will be at the cost of exposing the link to higher traffic. Let $\mathbf{R}_0 = (R_{0e}, e \in \mathcal{E})$, where $R_{0e} \leq 1$ is the *cut-off rate* on wireless link $e \in \mathcal{E}$, that is an upper bound for the rate R_e achievable with certain codes (e.g., convolutional codes) [21]. In general, the cut-off rate R_{0e} depends on the received signal-to-noise-ratio (SNR) and the modulation scheme being used. Given coding rate $R_e \leq R_{0e}$, the error probability on link e can be modeled as [21]

$$P_e = 2^{-T(R_{0e} - R_e)}, \quad (2)$$

where T is the coding block length. Based on the link failure model in (2), the probability that a packet is successfully transmitted along the k^{th} routing path, $k = 1, \dots, K_i$ for session $i \in \mathcal{I}$ is given by $\prod_{e \in \mathcal{E}, a_i^{ek}=1} (1 - P_e) = \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e)$. From the above equation, for each session $i \in \mathcal{I}$, the aggregate *effective* throughput at destination t_i becomes

$$\sum_{k=1}^{K_i} \alpha_i^k \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e). \quad (3)$$

To obtain the average end-to-end delay, we model each

link as a single $M/D/1$ queue based on the Kleinrock independence approximation [19]. Here, we assume that the arrival rates in the source nodes follow a Poisson distribution. Since the transmission times over all links are deterministic, the number of arrivals for each queue in any time interval can be assumed to follow a Poisson distribution with rate

$$\lambda_e = \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k / L, \quad (4)$$

where L is the packet length. From Little's Theorem [19], the average waiting time for each queue e corresponding to link $e \in \mathcal{E}$ is given by

$$\delta_e^Q = \frac{L \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'}}{2c_e R_e (c_e R_e - \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'})}, \quad (5)$$

where c_e denotes the *nominal* data rate of link $e \in \mathcal{E}$. By adding the waiting time and the transmission time $\delta_e^T = \frac{L}{c_e R_e}$ together, we have $\delta_e = \delta_e^Q + \delta_e^T$ for each link $e \in \mathcal{E}$. Then, the average end-to-end delay for each path $k = 1, \dots, K_i$ of session $i \in \mathcal{I}$ can be written as

$$\delta_i^k = \frac{L}{2} \sum_{e \in \mathcal{E}} \frac{a_e^{ik}}{c_e R_e} \left(2 + \frac{\sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'}}{c_e R_e - \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'}} \right). \quad (6)$$

To model the mutual interference among the wireless links in the network, we use the concept of *contention graph*. In the contention graph $G_C(\mathcal{V}_C, \mathcal{E}_C)$ corresponding to network $\mathcal{G}(\mathcal{V}, \mathcal{E})$, the set of vertices $\mathcal{V}_C$ represents the set of all wireless links $\mathcal{E}$ in the network graph G . An edge connects any two vertices in set $\mathcal{V}_C$ if the corresponding wireless links in the network graph mutually interfere with each other. That is, if the receiver node of one link is within the interference range of the sender node of the other link. Given the contention graph, each *complete* subgraph is called a *clique*. A *maximal clique* is a clique which is *not* a subgraph of any other clique [22]. We denote the set of all maximal cliques in G_C by $\mathcal{Q}$. In each instant, only one link among all members of a maximal clique $Q \in \mathcal{Q}$ can be active. The ratio $\frac{u_e}{c_e}$ denotes the portion of time that link $e \in \mathcal{E}$ is active when it is being used with data rate u_e . It is required that $\sum_{e \in Q} \frac{u_e}{c_e} \leq \nu$ for each clique $Q \in \mathcal{Q}$ where $\nu \in (0, 1]$ is called the *clique capacity*. Note that $\nu = 1$ is a necessary constraint on the clique capacity. It may not always be possible to find feasible schedules that achieve a clique capacity of $\nu = 1$. Shannon showed that $\nu = \frac{2}{3}$ is a sufficient condition on the clique capacity in order to obtain a feasible schedule for the links in the clique [23].

We formulate the problem of jointly allocating coding rates and sending data rates such that the utility of the network be maximized. The utility of each session $i \in \mathcal{I}$ is defined as

$$U_i(\alpha, R) = (1 - w) \sum_{k=1}^{K_i} \alpha_i^k \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e) - w \sum_{k=1}^{K_i} \delta_i^k, \quad (7)$$

where δ_i^k is as in (6). Here, the utility of session i is a *weighted trade-off* between the session's aggregate effective throughput and its average delay. It is a trade-off because it can

be increased by either increasing the throughput or decreasing the delay. We can tune the importance of delay by changing its weight, w . By increasing w , we move on the trade-off curve towards decreasing delay at the cost of decreasing the throughput. We define the utility of the network as either the summation of all utilities of data sessions $i \in \mathcal{I}$, or just the one with the minimum value.

1) *Maximizing the Aggregate Utility of the Network*: This problem is formulated as

$$\begin{aligned} & \underset{\alpha \geq 0, 0 < R \leq R_0}{\text{maximize}} && (1 - w) \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \alpha_i^k \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e) \\ & && - w \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \delta_i^k \\ & \text{subject to} && \sum_{e \in Q} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k R_e^{-1} c_e^{-1} \leq \nu, \quad Q \in \mathcal{Q}, \\ & && \delta_i^k \leq \delta_i^{max}, \quad i \in \mathcal{I}, k = 1, \dots, K_i, \end{aligned} \quad (8)$$

where δ_i^{max} is the maximum delay that can be tolerated for each path of session $i \in \mathcal{I}$. The set of constraints declare that the portion of time that all links in a maximal clique are active must be less than the clique capacity. The expressions for P_e and δ_i^k are as in (2) and (6), respectively.

2) *Maximizing the Minimum Utility of the Network*: This problem is formulated as

$$\begin{aligned} & \underset{\alpha \geq 0, 0 < R \leq R_0}{\text{maximize}} && \min_{i \in \mathcal{I}} (1 - w) \sum_{k=1}^{K_i} \alpha_i^k \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e) - w \sum_{k=1}^{K_i} \delta_i^k \\ & \text{subject to} && \sum_{e \in Q} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k R_e^{-1} c_e^{-1} \leq \nu, \quad Q \in \mathcal{Q}, \\ & && \delta_i^k \leq \delta_i^{max}, \quad i \in \mathcal{I}, k = 1, \dots, K_i. \end{aligned} \quad (9)$$

Unlike (8), here the design addresses the notion of maxmin fairness among sessions.

III. DELAY-AWARE OPTIMAL DATA RATE AND CODING RATE ALLOCATION

The optimization problem in (8) is non-convex and non-separable due to the product forms in the objective function with respect to the effective throughput, the fractional forms in the first set of constraints and in the delay constraints in (6), the exponential forms in the objective function with respect to error probabilities, the non-separability of reliability and throughput due to multi-path routing, and the coupling across variables because of delay constraints and channel coding. Most of the above properties are due to the fact that we consider multi-path routing and wireless interference. For example, if we assume there is no interference, which is true for wired networks, the clique capacity constraints would reduce to *linear* link capacity constraints for any link $e \in \mathcal{E}$:

$$\frac{1}{R_e c_e} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k \leq 1, \quad \Rightarrow \quad \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k - R_e c_e \leq 0. \quad (10)$$

We can also show that the non-convexity due to the product forms in the objective function can be resolved if there is

only a *single* routing path for each session [6]. However, all sources of complexity remain in place when multi-path routing is used and wireless transmissions are subject to interference. In the following, we use various techniques to overcome the complexity of the problem formulation and convert problem (8) into a convex problem.

Consider the exponential form of P_e in (2). For notational simplicity, we can rewrite (2) as $P_e = X_e \exp(Z R_e)$ for each link $e \in \mathcal{E}$, where $X_e = 2^{-TR_0}$ and $Z = T \ln 2$. We can use *Taylor series expansion* and rewrite the above equation as $P_e = X_e \sum_{n=0}^{\infty} \frac{(Z R_e)^n}{n!}$. Clearly, for some *bounded* integer $N_e \gg 1$, we can approximate P_e as $X_e \sum_{n=0}^{N_e} \frac{(Z R_e)^n}{n!}$ for each link $e \in \mathcal{E}$. We investigate the value of N_e necessary for obtaining a good approximation through simulation. If the error probabilities P_e are small, we can rewrite the receiving rates in each session as

$$\sum_{k=1}^{K_i} \alpha_i^k \prod_{e \in \mathcal{E}} (1 - a_i^{ek} P_e) \approx \sum_{k=1}^{K_i} \alpha_i^k \left(1 - \sum_{e \in \mathcal{E}} a_i^{ek} P_e \right). \quad (11)$$

Due to the polynomial forms in the objective function and the constraints, we can solve problem (8) by using geometric programming techniques. In this respect, using the approximated value for P_e , we replace (11) in the objective function of problem (8) and introduce variable t such that t is a lower-bound for the objective function. That is,

$$\begin{aligned} & t + (1-w) \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \sum_{e \in \mathcal{E}} \sum_{n=0}^{N_e} \frac{\alpha_i^k a_i^{ek} X_e (Z R_e)^n}{n!} + w \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \delta_i^k \\ & \leq (1-w) \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \alpha_i^k. \end{aligned} \quad (12)$$

Then, we follow the signomial programming techniques [20] to approximate the polynomial in the right-hand side of (12), which is only a function of α , as a *monomial*, i.e., a polynomial with only *one* term and *positive* multiplier. This approximation can be done around some initial point $\hat{\alpha}$. For a parameter $f_s > 1$, which is close to 1, we have

$$\begin{aligned} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \alpha_i^k & \approx \left(\sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \hat{\alpha}_i^k \right) \prod_{i \in \mathcal{I}} \prod_{k=1}^{K_i} \left(\frac{\alpha_i^k}{\hat{\alpha}_i^k} \right)^{\hat{\alpha}_i^k / \left(\sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} \hat{\alpha}_{i'}^{k'} \right)}, \\ & \forall \alpha \in [\hat{\alpha}/f_s, f_s \hat{\alpha}], \end{aligned} \quad (13)$$

where $[\hat{\alpha}/f_s, f_s \hat{\alpha}]$ is a small neighborhood around initial point $\hat{\alpha}$.

For notational convenience, we define $\hat{\Lambda}$, which only depends on the initial point $\hat{\alpha}$, as $\hat{\Lambda}^{-1} = \left(\sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \hat{\alpha}_i^k \right)$. Then, inequality (12) can be approximated around the initial point $\hat{\alpha}$ as

$$\begin{aligned} & \frac{\hat{\Lambda}}{1-w} \left(t + (1-w) \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \sum_{e \in \mathcal{E}} \sum_{n=0}^{N_e} \frac{\alpha_i^k a_i^{ek} X_e (Z R_e)^n}{n!} \right. \\ & \quad \left. + w \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \delta_i^k \right) \prod_{i \in \mathcal{I}} \prod_{k=1}^{K_i} \left(\frac{\alpha_i^k}{\hat{\alpha}_i^k} \right)^{-\hat{\alpha}_i^k \hat{\Lambda}} \leq 1. \end{aligned} \quad (14)$$

The above constraint is a *posynomial*, i.e., a polynomial with

only *positive* terms. Posynomials are the building blocks in geometric programming [20]. By minimizing t^{-1} , we maximize the objective function in (8).

To tackle the fractional forms in the delay constraints, we can write (6) in an inequality form

$$\delta_i^k \geq \frac{L}{2} \sum_{e \in \mathcal{E}} \frac{a_i^{ek}}{c_e R_e} \left(2 + \frac{\sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'}}{\left(c_e R_e - \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'} \right)} \right), \quad i \in \mathcal{I}, k = 1, \dots, K_i. \quad (15)$$

It can be shown that (15) is always satisfied with equality for the optimal solution. This can be proved by contradiction. Assume that (15) is not satisfied with equality in the *optimal solution* for some i and k . Note that δ_i^k is not lower bounded in any other set of constraints. Therefore, we can decrease δ_i^k such that the corresponding constraint is satisfied with equality. This leads to further increasing the value of the objective function by choosing a solution different than the optimal solution which is a contradiction. That is, it is an active inequality constraint. For each link $e \in \mathcal{E}$, we introduce new variables Y_e such that

$$\frac{Y_e^{-1}}{c_e} + \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \frac{\alpha_{i'}^{k'} R_e^{-1}}{c_e} \leq 1, \quad \forall e \in \mathcal{E}. \quad (16)$$

We can show that (15) is satisfied if (16) holds and we have

$$\delta_i^k \geq \frac{L}{2} \sum_{e \in \mathcal{E}} a_i^{ek} \left(\frac{2}{c_e R_e} + \frac{Y_e \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'}}{c_e R_e^2} \right), \quad \forall i \in \mathcal{I}, k = 1, \dots, K_i. \quad (17)$$

Similarly, we can show that (16) and (17) are always satisfied with equality for the optimal solution.

By introducing t as in (14) and adding constraints (16), and (17) to the constraints of problem (8), it is equivalent to the problem (18) in which the objective is to minimize t^{-1} which is equal to maximizing t where t is declared to be a lower bound for the network utility function in the first set of constraints. The second set of constraints declare the capacity constraints and the third and fourth set are active constraints (17) and (16), respectively. The last set of constraints guarantees all end-to-end delays to be bounded:

$$\begin{aligned} & \underset{t > 0, \hat{\alpha}/f_s \leq \alpha \leq f_s \hat{\alpha}, 0 < R \leq R_0, \delta > 0, Y > 0}{\text{minimize}} \quad t^{-1} \\ & \text{subject to} \\ & \frac{\hat{\Lambda}}{1-w} \left(t + (1-w) \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \sum_{e \in \mathcal{E}} \sum_{n=0}^{N_e} \frac{\alpha_i^k a_i^{ek} X_e (Z R_e)^n}{n!} \right. \\ & \quad \left. + w \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \delta_i^k \right) \prod_{i \in \mathcal{I}} \prod_{k=1}^{K_i} \left(\frac{\alpha_i^k}{\hat{\alpha}_i^k} \right)^{-\hat{\alpha}_i^k \hat{\Lambda}} \leq 1, \\ & \frac{1}{\nu} \sum_{e \in \mathcal{Q}} \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} a_i^{ek} \alpha_i^k R_e^{-1} c_e^{-1} \leq 1, \quad \forall Q \in \mathcal{Q}, \end{aligned}$$

$$\begin{aligned} \frac{L}{2} \sum_{e \in \mathcal{E}} \frac{a_i^{ek}}{c_e} \left(2R_e^{-1} + R_e^{-2} Y_e \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} \alpha_{i'}^{k'} \right) \delta_i^{k-1} &\leq 1, \\ \forall i \in \mathcal{I}, k = 1, \dots, K_i, \\ \frac{Y_e^{-1}}{c_e} + \sum_{i' \in \mathcal{I}} \sum_{k'=1}^{K_{i'}} a_{i'}^{ek'} R_e^{-1} \alpha_{i'}^{k'} &\leq 1, \quad \forall e \in \mathcal{E}, \\ \delta_i^k &\leq \delta_i^{max}, \quad \forall i \in \mathcal{I}, k = 1, \dots, K_i. \end{aligned} \quad (18)$$

Problem (18) is a standard geometric programming problem that can be converted into a convex optimization problem [20] and can be solved around an initial point. It has been shown that iteratively solving (18) converges to the optimal solution of problem (8) [20]. In each iteration, (18) is initialized with the optimal solution of the problem corresponding to the last iteration. As discussed, we may move on the throughput-delay trade-off curve by tuning the delay importance weight w . w must be chosen such that the objective function in problem (8) remains positive.

Similarly, we can convert problem (9) into a convex optimization problem. Compared to solving problem (8), there are two differences. First, variable t is introduced such that

$$t \leq (1-w) \sum_{k=1}^{K_i} \alpha_i^k \left(1 - \sum_{e \in \mathcal{E}} a_i^{ek} X_e \sum_{n=0}^{N_e} \frac{(Z R_e)^n}{n!} \right) - w \sum_{k=1}^{K_i} \delta_i^k, \quad \forall i \in \mathcal{I}. \quad (19)$$

As in the earlier case, we use signomial techniques to convert (19) into a constraint in the standard form of geometric programming problems. By applying a similar technique in (13) this time to $\sum_{k=1}^{K_i} \alpha_i^k$, we can rewrite (19) as

$$\begin{aligned} \frac{\hat{\Lambda}_i}{1-w} \left(t + (1-w) \sum_{k=1}^{K_i} \sum_{e \in \mathcal{E}} \sum_{n=0}^{N_e} \frac{\alpha_i^k a_i^{ek} X_e (Z R_e)^n}{n!} + w \sum_{k=1}^{K_i} \delta_i^k \right) \\ \times \prod_{k=1}^{K_i} \left(\frac{\alpha_i^k}{\hat{\alpha}_i^k} \right)^{-\hat{\alpha}_i^k \hat{\Lambda}_i} \leq 1, \quad \forall i \in \mathcal{I}, \end{aligned} \quad (20)$$

where $\hat{\Lambda}_i^{-1} = (\sum_{k=1}^{K_i} \hat{\alpha}_i^k)$. Second, inequalities (16) and (17) may not be active anymore and so inequality (15) may not be satisfied with equality. In this case, we are in fact dealing with the upper bounds of the average end-to-end delay in the objective function. In this way, the performance of the network is better than what we would expect from the obtained solution in terms of average delay since the upper bounds on the average delay are used in the objective function. The rest of the formulation is the same as the one in (18). Again, w must be chosen such that the objective function in problem (9) remains positive.

IV. NUMERICAL RESULTS

In this section, we numerically solve problem (8) to determine the optimal sending and coding rates such that the utility of the network (i.e., the trade-off between the aggregate effective throughput and the average delay) is maximized. We show that maximizing the aggregate network utility leads to higher benefits in the throughput-delay trade-off (i.e., we obtain lower delays at the cost of lower decrease in throughput),

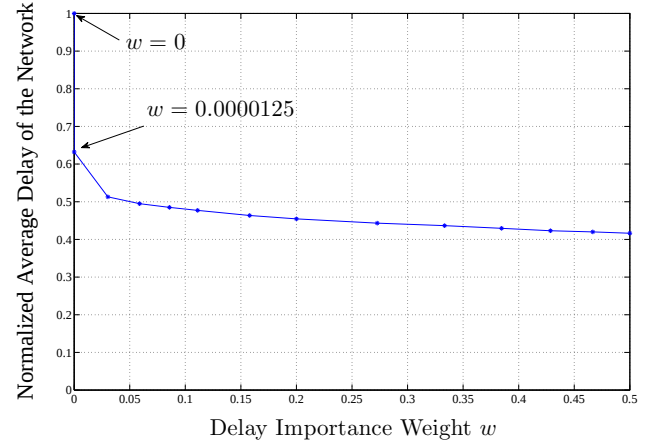


Fig. 2. The average delay decreases as its importance weight in the objective function increases. Delay values are normalized over the corresponding values for which $w = 0$. We observe a decrease of almost 58% when $w = 0.5$.

compared to the case of maximizing the minimum utility in the network. In the former case, no fairness is achieved and some sessions may be starved. Therefore, we also solve problem (9) to determine the sending rates and coding rates such that the minimum utility of the network is maximized. We show that at the cost of loosing some gain in the throughput-delay trade-off we can provide fairness among all sessions. In our set of simulations, we use $T = 10$, $L = 8000$ bits, $c_e = 11$ Mbps, $N_e = 15$, $f_s = 1.1$, $R_{0e} = 1$, and $\nu = \frac{2}{3}$ [23]. The proper values for parameters N_e and f_s are obtained from simulations.

We solve problem (8) for different values of w in a feasible range across 50 different random topologies. Each random topology is a 5×5 grid topology for which 20 nodes are placed in random locations. Five pairs of nodes are randomly selected as the source and destination pairs. We observe that by considering the delay in the objective function, even with a small weight, we can decrease the average delay by 37% compared to the case with no delay consideration (Fig. 2). We also note that by further increasing the importance weight of delay in the objective function the average delay decreases by 58%. The more a link is utilized, the higher will be the queuing delays for that link. Therefore, decreasing the average delay leads to the use of the intermediate links at a lower rate that in turn leads to a slight decrease (0.1%) in aggregate throughput. This confirms the delay-throughput trade-off. We can also see that by decreasing the throughput by a small amount, the delay decreases dramatically at the starting point. This is because delay is an exponential function of the utilization rate.

Maximizing the total throughput usually requires sacrificing fairness among sessions. That is, some sessions may starve while some other sessions use the network with a higher throughput. For instance after solving problem (8) for a sample topology, we observe that sessions 1 and 3 use the network at a rate of 2 Mbps while sessions 2 and 5 starve and session 4 sends data at a rate of 4.5 Mbps. To provide fairness among sessions, we solve problem (9) to maximize the minimum utility across sessions for different feasible choices of parameter w and for 50 randomly selected topologies. Maximizing the

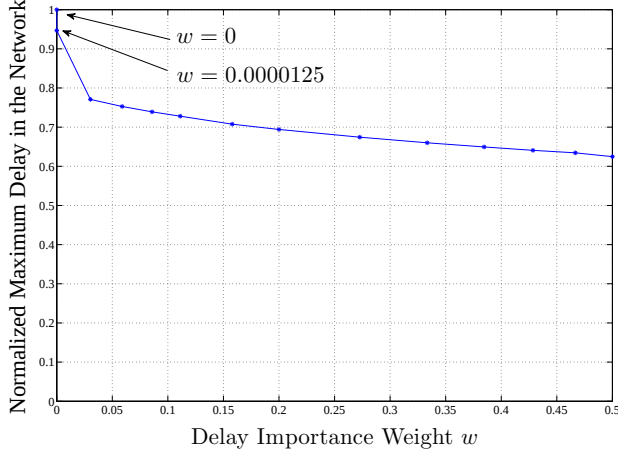


Fig. 3. The normalized maximum delay among the users in the network decreases as the corresponding importance weight in the objective function increases. Delay values are normalized over the values corresponding to $w = 0$. We observe almost 37% decrease in the maximum delay in the network when w reaches 0.5.

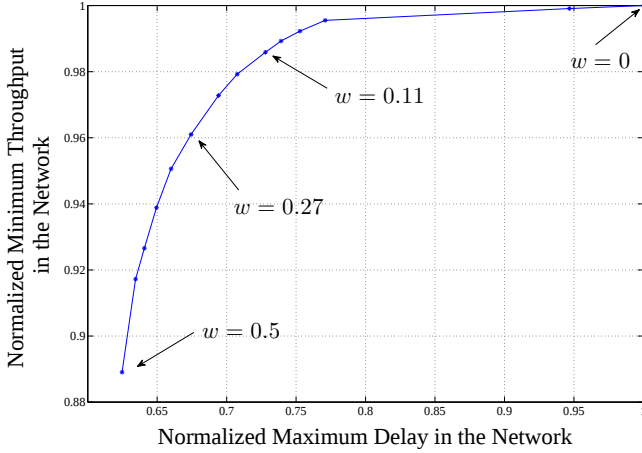


Fig. 4. The trade-off between the maximum delay and the minimum throughput in the network. We may move on the curve by tuning w , the delay importance weight. Delay and throughput are normalized over their corresponding values at $w = 0$.

minimum utility of the network, we expect that there will be no session with starvation. We can see that the normalized maximum end-to-end delay in the network among all routing paths decreases by almost 5% on average when we consider the delay only by a very small weight (Fig. 3). We can further decrease the delay by 37% when w increases. Again, there exists a trade-off between the maximum end-to-end delay of the network and the aggregate throughput of the session with the minimum aggregate throughput in Fig. 4. By tracing the graph starting from the upper right corner ($w = 0$) to the lower-left corner ($w = 0.5$), we gain a 37% improvement in the maximum delay at a cost of a slight decrease at only 11% in the minimum throughput when we are in proper points of the curve.

We use Jain's fairness index to quantitatively measure the fairness of the throughput attained among different unicast sessions. Let Ψ denote Jain's fairness index. We have $\Psi =$

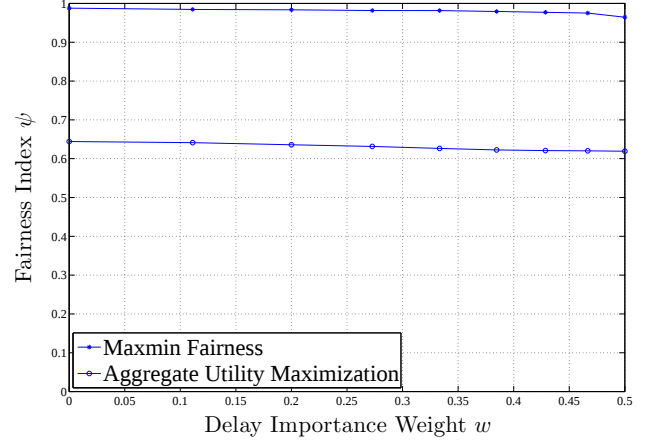


Fig. 5. Fairness index is shown for both problem (8) (aggregate utility maximization) and problem (9) (maxmin fairness).

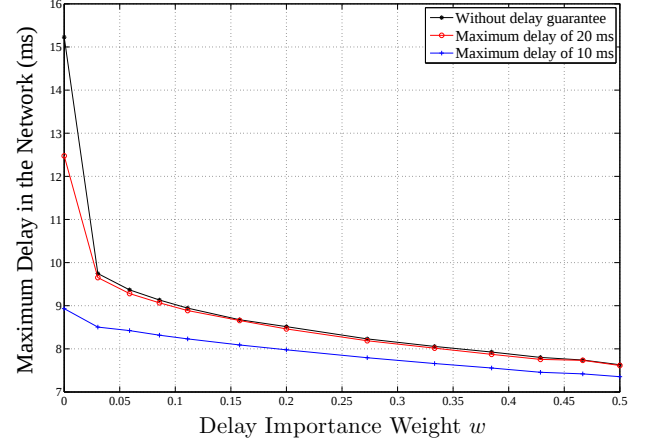


Fig. 6. Average maximum delay is shown when there is no guarantee on the maximum delay, there is a guarantee of 10 ms, and 20 ms. Since the guarantee is imposed on the upper bound of the delay, the delay may not reach the guaranteed value.

$\frac{(\sum_{i \in \mathcal{I}} x_i)^2}{|\mathcal{I}| \sum_{i \in \mathcal{I}} x_i^2}$ where x_i denotes the total effective throughput of flow $i \in \mathcal{I}$ from (3). We can see in Fig. 5 that fairness is improved when the resource allocation is based on the solution of problem (9).

By now, we considered the normalized values of delay over their value when $w = 0$ (delay is not considered). It is interesting to see how the average maximum delays in the network change under different delay guarantees. We solve problem (9) without considering the last set of constraints (delay guarantee constraints), and also when δ_i^{max} is 10 ms and 20 ms for all $i \in \mathcal{I}$ (Fig. 6). We can see that the delay is guaranteed to be less than δ_i^{max} .

As mentioned earlier, the decrease in delay is gained at the cost of decreasing the utility of the links which in turn leads to a decrease in the overall throughput. This throughput degradation can be compensated for by using channel coding. To determine the effect of channel coding, we consider the throughput-delay trade-off in a network in which channel coding is *not performed* and we study how this can affect the network performance. We assume a packet error rate of 30%

at each link and solve problem (9) without channel coding to show how it affects the performance. By increasing the weight of delay, we can only decrease the maximum delay by around 3% at the cost of 22% decrease in minimum throughput in Table I. w varies in its feasible range. This shows how performance degrades when channel coding is not used and reveals the importance of channel coding.

TABLE I

THE TRADE-OFF BETWEEN THE MAXIMUM DELAY AND THE MINIMUM THROUGHPUT IN THE NETWORK WHEN CHANNEL CODING IS NOT BEING USED.

w	Normalized Maximum Delay	Normalized Minimum Throughput
0	1	1
0.3	0.97	0.78

V. CONCLUSION

In this paper, we considered the trade-off between decreasing the end-to-end delay and increasing the aggregate throughput in wireless networks with channel coding and showed that a noticeable enhancement across both design goals is feasible if a combination of multi-path routing and adaptive channel coding are employed. We jointly formulated the end-to-end data rate allocation and adaptive channel coding within the general framework of network utility maximization (NUM) under two variations. The first problem is formulated for maximization of the aggregate network utility, i.e., the overall system performance. The second problem is formulated for maximization of the minimum utility among the end-to-end flows to achieve fairness. Due to non-convexities such as in product terms and fractional terms in the objective function and the constraints, the formulated optimization problems are non-convex and non-separable, and difficult to solve. Nevertheless, we introduced an algorithm that can solve the two NUM problems with low computational complexity. Through our simulation studies, we note that, in many cases, significant improvement in end-to-end delay can be obtained with marginal decrease in aggregate throughput, suggesting that satisfying stringent delay requirements can be achieved if multi-path routing and adaptive channel coding are employed. The fair resource allocation aspect of our proposed design is of interest in vehicular networks where multiple vehicles share an access point in order to obtain connectivity to the Internet. The centralized solution that we have proposed in this paper can particularly be used in the case when the stationary access point provides connectivity for all vehicles that it serves. It can also serve as a benchmark for distributed algorithms which are to be developed in future. Nevertheless, a distributed algorithm can support much broader ranges of application types. For future work, we plan to study the possibility of finding the data and channel code rates in a distributed manner using Lyapunov stability theory, similar to the backpressure algorithms [24].

REFERENCES

- [1] D. Ren, Y. T. H. Li, and S. H. G. Chan, "Fast-mesh: A low-delay high-bandwidth mesh for peer-to-peer live streaming," *IEEE Trans. on Multimedia*, vol. 11, no. 8, pp. 1446–1456, Dec. 2009.
- [2] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of Operations Research Society*, vol. 49, no. 3, pp. 237–252, Mar. 1998.
- [3] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.
- [4] L. Chen, S. Low, and J. Doyle, "Joint congestion control and media access control design for ad hoc wireless networks," in *Proc. of IEEE Infocom*, Miami, FL, Mar. 2005.
- [5] D. O'Neill, B. S. Thian, A. Goldsmith, and S. Boyd, "Wireless NUM: Rate and reliability tradeoffs in random environment," in *Proc. of IEEE WCNC*, Budapest, Hungary, Apr. 2009.
- [6] J. W. Lee, M. Chiang, and A. Calderbank, "Price-based distributed algorithms for rate-reliability based tradeoff in network utility maximization," *IEEE J. Select. Areas Commun.*, vol. 24, no. 5, pp. 962–976, May 2006.
- [7] H. Ma, D. Fayek, and P. H. Ho, "Availability-constrained multipath protection in backbone networks with double-link failure," in *Proc. of IEEE ICC*, Beijing, China, May 2008.
- [8] T. Cinkler and L. Gyarmati, "MPP: Optimal multi-path routing with protection," in *Proc. of IEEE ICC*, Beijing, China, May 2008.
- [9] B. Yahya and J. B. Othman, "REER: Robust and energy efficient multipath routing protocol for wireless sensor networks," in *Proc. of IEEE Globecom*, Honolulu, HI, Dec. 2009.
- [10] Y. Li, M. Chiang, A. R. Calderbank, and S. N. Diggavi, "Optimal rate-reliability-delay tradeoff in networks with composite links," in *Proc. of IEEE Infocom*, Anchorage, Alaska, May 2007.
- [11] K. Ronasi, H. Mohsenian-Rad, V. W. S. Wong, S. Gopalakrishnan, and R. Schober, "Optimal data transmission and channel code rate allocation in multi-path wireless networks," *Submitted to IEEE ICC*, Kyoto, Japan, 2011.
- [12] D. O'Neill, A. J. Goldsmith, and S. Boyd, "Cross-layer design with adaptive modulation: Delay, rate, and energy tradeoffs," in *Proc. of IEEE Globecom*, New Orleans, LA, Dec. 2008.
- [13] M. Saad, A. Leon-Garcia, and W. Yu, "Optimal network rate allocation under end-to-end quality-of-service requirements," *IEEE Trans. on Network and Service Management*, vol. 4, no. 3, pp. 40–49, Dec. 2007.
- [14] —, "Delay constrained optimal resource utilization of wireless networks for distributed control systems," *IEEE Communications Letters*, vol. 12, no. 4, pp. 289–291, Apr. 2008.
- [15] M. G. Kallitsis, R. D. Callaway, M. Devetsikiotis, and G. Michailidis, "Distributed and dynamic resource allocation for delay sensitive network services," in *Proc. of IEEE Globecom*, New Orleans, LA, Dec. 2008.
- [16] Y. Li, M. Chiang, and A. R. Calderbank, "Congestion control in networks with delay sensitive traffic," in *Proc. of IEEE Globecom*, Washington, DC, Nov. 2007.
- [17] S. Kalyanasundaram, E. K. P. Chong, and N. B. Shroff, "Optimal resource allocation in multi-class networks with user-specified utility functions," *Comput. Networks*, vol. 38, no. 5, pp. 613–630, Apr. 2002.
- [18] Y. Li, M. Chiang, R. Calderbank, and S. N. Diggavi, "Optimal rate-reliability-delay tradeoff in networks with composite links," *IEEE Trans. on Communications*, vol. 57, no. 5, pp. 1390–1401, May 2009.
- [19] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 2nd Edition, 1992.
- [20] M. Chiang, *Geometric Programming for Communication Systems*. Foundations and Trends in Communications and Information Theory, 2005.
- [21] J. G. Proakis, *Digital Communications*. 4th edition, New York: McGraw-Hill, 2001.
- [22] R. Gupta and J. Walrand, "Approximating maximal cliques in ad-hoc networks," in *Proc. of IEEE PIMRC*, Barcelona, Spain, Sept. 2004.
- [23] C. Shannon, "A theorem on coloring the lines in the network," *J. Math. Phys.*, vol. 28, pp. 148–151, Sept. 1949.
- [24] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource Allocation and Cross Layer Control in Wireless Networks*. Foundations and Trends in Networking, 2006.

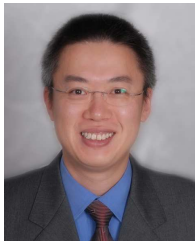


Keivan Ronasi (S'05) was born in Isfahan, Iran in 1981. He received the B.Sc. degree in electrical and computer engineering from Isfahan University of Technology, Isfahan, Iran in 2004. He received the M.Sc. degree from the University of Tehran, Tehran, Iran in 2007. Currently, he is a Ph.D. candidate at the University of British Columbia, Vancouver, Canada. His research interests include modeling and design of reliable wireless communication networks, network coding, and stochastic network optimization.



Amir-Hamed Mohsenian-Rad (S'04, M'09) received masters degree in Electrical Engineering from Sharif University of Technology in 2004 and Ph.D. degree in Electrical and Computer Engineering from The University of British Columbia (UBC) in 2008. Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering at Texas Tech University. Dr. Mohsenian-Rad is the recipient of the UBC Graduate Fellowship, Pacific Century Graduate Scholarship, and the Natural Sciences and Engineering Research Council of Canada

(NSERC) Post-doctoral Fellowship. He is an Associate Editor of the Elsevier International Journal of Electronics and Communication and serves as a Technical Program Committee member in various conferences, including IEEE Globecom, ICC, VTC, and CCNC. His research interests include the design, optimization, and game-theoretic analysis of computer and communication networks and smart power grids.



Vincent W.S. Wong (SM'07) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently an Associate Professor. His research areas include protocol design,

optimization, and resource management of communication networks, with applications to the Internet, wireless networks, smart grid, RFID systems, and intelligent transportation systems. Dr. Wong is an Associate Editor of the IEEE Transactions on Vehicular Technology and an Editor of KICS/IEEE Journal of Communications and Networks. He is the Symposium Co-chair of IEEE Globecom'10, Wireless Communications Symposium. He serves as TPC member in various conferences, including IEEE Infocom and ICC.



Sathish Gopalakrishnan is an Assistant Professor of Electrical and Computer Engineering at the University of British Columbia. His research interests center around resource allocation problems in several contexts including real-time, embedded systems and wireless networks. Prior to joining UBC in 2007, he obtained a Ph.D. in Computer Science and an M.S. in Applied Mathematics from the University of Illinois at Urbana-Champaign. He has received awards for his work from the IEEE Industrial Electronics Society (Best Paper in the IEEE Transactions

on Industrial Informatics in 2008) and at the IEEE Real-Time Systems Symposium (in 2004).



Robert Schober (S'98, M'01, SM'08, F'10) was born in Neuendettelsau, Germany, in 1971. He received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen-Nuermberg in 1997 and 2000, respectively. From May 2001 to April 2002 he was a Postdoctoral Fellow at the University of Toronto, Canada, sponsored by the German Academic Exchange Service (DAAD). Since May 2002 he has been with the University of British Columbia (UBC), Vancouver, Canada, where he is now a Full Professor and

Canada Research Chair (Tier II) in Wireless Communications. His research interests fall into the broad areas of Communication Theory, Wireless Communications, and Statistical Signal Processing.

Dr. Schober received the 2002 Heinz MaierLeibnitz Award of the German Science Foundation (DFG), the 2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, and the 2008 Charles McDowell Award for Excellence in Research from UBC. In addition, he received best paper awards from the German Information Technology Society (ITG), the European Association for Signal, Speech and Image Processing (EURASIP), IEEE ICUWB 2006, the International Zurich Seminar on Broadband Communications, and European Wireless 2000. Dr. Schober is also the Area Editor for Modulation and Signal Design for the IEEE Transactions on Communications.