Multi-rate Multicasting with Network Coding

(Invited Paper)

Subhash Lakshminarayana
Electrical and Computer Engineering
The Ohio State University
2015 Neil Avenue, Columbus, OH 43210
{lakshminarayana.1@osu.edu}

Atilla Eryilmaz[*]
Electrical and Computer Engineering
The Ohio State University
2015 Neil Avenue, Columbus, OH 43210
{eryilmaz@ece.osu.edu}

ABSTRACT
Single-rate multicasting may yield low utilization of the network resources when a subset of the receivers creates a bottleneck for the whole multicast group. Thus, practical multi-rate multicasting strategies are required to allow users with better channels to achieve maximum performance. While earlier studies have investigated such mechanisms using routing strategies over fixed trees, in this work we consider the possibility of using network coding without a given subgraph. Our scheme identifies the optimal routes and provides the utility maximizing rate allocation and coding solution. We also propose and investigate a *nested multicasting method* to reduce the complexity of the original scheme for practical implementation.

Categories and Subject Descriptors
C.2 [**Network Architecture and Design**]: Distributed Networks, Network Communications; G.1.6 [**Optimization**]: Convex Programming; F.2 [**Analysis of Algorithms and Complexity**]: General

Keywords
Multi-rate Multicast, Network Coding, Convex Optimization, Primal-dual method, Congestion Control, Utility Maximization

1. INTRODUCTION
Multi-hop communication networks are expected to efficiently serve many applications with diverse characteristics and quality-of-service constraints. Many of these applications, including voice/video broadcast and file sharing, generate data to be

multicast to a group of destinations rather than a single one. Traditionally, the rate of such a *single-rate* multicast session is chosen such that all the receivers can successfully receive the transmitted information at the selected rate. It is shown in [1] and [15] that *linear network coding* can be used in such a scenario to achieve the maximum possible allowable rate. With this strategy, nodes are allowed to "mix" the packets of the session as they traverse them, and thus can improve the achievable rate of the session. Many works have exploited such advantages (e.g. [6, 11, 7, 4]).

However, the single-rate nature of the traditional multicasting strategy is limiting because its throughput is restricted by the capacity to the *bottleneck destination*. In other words, all the destinations suffer from the constraints associated with the destination under the worst conditions. In order to overcome this problem, we study *multi-rate* multicast capabilities, where the source is allowed to multicast its data to different destinations at different rates based on the condition of the network to them. With this capability, the bottleneck destinations will not able to throttle the whole communication quality.

There are a number of papers focusing on multi-rate multicasting scenario, but without the capability of network coding. In particular, Kar et.al. [13] propose an optimization-based approach for a given fixed routing tree associated with each multicast session. In another related work, Bui et.al. [3] propose multi-rate multicasting based on scheduling virtual (shadow) "traffic" that "moves" in reverse direction from destinations to sources. This work also assumes a single fixed tree for routing the traffic and network coding is not utilized. Sarkar et. al. [22] propose multicasting with M-Best trees and switching between the trees depending on congestion level. Switching between trees gives some sort of dynamism to the algorithm. But the M-Best trees are still assumed to be pre-defined.

In this paper, we propose a distributed rate allocation, scheduling, routing, and coding solution to multi-rate multicasting for general multihop networks when network coding is allowed, and when the routes are not predefined. The main novelties of this paper are:

• We formulate the multi-rate multicast utility maximization problem by decomposing the multi-rate session into many single-rate subsessions, and optimizing over the subsession rates.

• We do not assume presence of pre-defined routes or trees. The algorithm automatically forms the optimal subgraphs in addition to finding the optimal rates.

• The algorithm makes provision for intra-session network coding, which enables higher network throughput.

• We propose a novel extension to the original algorithm to reduce the complexity of the implementation significantly. This low-complexity algorithm utilizes the special structure of the multi-rate multicast problem to construct the optimal subsessions gradually rather than optimizing over an exponential number of them.

The rest of the paper is organized as follows. Section 2 introduces the system model and sets up the multi-rate multicast problem. Then, Section 3 provides the design of our cross-layer mechanism through the use of dual methods. In Section 4, we provide the analysis of the proposed mechanism and its optimality characteristics. In Section 5, we exploit the special structure of the problem to provide a low-complexity implementation for the cross-layer mechanism. Section 6 includes extensive simulation results for both the original and low-complexity algorithms. Our concluding remarks are provided in Section 7.

2. MODEL AND PROBLEM DESCRIPTION

Consider a fixed multihop network that is described by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes with cardinality N and $\mathcal{L}$ is the set of directed links. While in a wireless network, the achievable link rates are interdependent due to interference effects, in a wired network it can be assumed that each link (i, j) has a fixed capacity of $c_{ij} \in \{0, 1, 2, \cdots\}$. Although our results can be extended to cover wireless conditions, in this paper we will focus on wired networks in order to avoid unnecessary complications and to provide the main idea more clearly.

We denote the *multi-rate multicast session* with rate $\mathbf{y} := (y_d)_{d \in \mathcal{D}}$ from source node s to the set of destination nodes $\mathcal{D}$ by $(s, \mathcal{D}, \mathbf{y})$. The multi-rate nature of the multicast transmission allows destination d to be served at rate y_d, subject to achievability constraints. Throughout the paper, we will focus on the scenario of a single multi-rate multicast session for ease of discussion. However, we note that all of our results can be extended to multiple multi-rate multicast sessions when network coding is only allowed within a session.

We decompose multi-rate multicast session $(s, \mathcal{D}, \mathbf{y})$ into many single-rate multicast *subsessions*, where each subsession, $(s, \mathcal{D}_k, x_k)$ is described by the source node s, a subset of destinations $\mathcal{D}_k \subset \mathcal{D}$, and a single-rate x_k. Accordingly, the potential number of subsessions of a given session is $K := 2^{|\mathcal{D}|}$. Note that, with such a decomposition, the destination rates $\mathbf{y}$ and the subsession rates $\mathbf{x} := (x_k)_{k=1}^{K}$ are related as

$$y_d(\mathbf{x}) = \sum_{k \ni d} x_k, \qquad \text{for all } d \in \mathcal{D}, \qquad (1)$$

where $k \ni d$ means $\{k : d \in \mathcal{D}_k\}$, i.e., all subsessions that have d as a destination.

Subsession k injects packets with an average rate of x_k to be multicast to its destination group $\mathcal{D}_k$, where each packet is represented by a vector over a finite field $\mathbb{F}_r$, where r is assumed to be a large positive integer. In our framework, we allow each subsession to perform *intra-session network coding*. Specifically, as the packets of the same subsession traverse the network nodes, we allow the node to create random linear combinations of the packets before transmitting them to the neighboring nodes. In particular, a random linear combination of the packets $\{\mathbf{P}_1, \cdots, \mathbf{P}_J\}$ is given by $\sum_{j=1}^{J} \theta_j \mathbf{P}_j$, where $\{\theta_j\}$ are randomly selected coefficients from the finite field. Such intra-session coding operations are known to improve the achievable throughput region of the network (e.g. [1, 12]). We aim to exploit such network coding capabilities for subsessions while optimizing the subsession rates to yield the best multi-rate session performance.

We assume that the utilization achieved at each destination, say $d \in \mathcal{D}$, is measured by a utility function $U_d(\cdot)$ of its rate y_d. We make the following assumptions on the utility function:

• $U_d(.)$ is a strictly concave, twice differentiable, non-decreasing function of mean rate y_d.

• For every m and M satisfying $0 < m < M < \infty$, there exists constants $\tilde{c}$ and $\tilde{C}$ satisfying $0 < \tilde{c} < \tilde{C} < \infty$ such that

$$\tilde{c} \le -\frac{1}{U_d''(y_d)} \le \tilde{C}, \quad \forall y_d \in [Km, KM],$$

where $K = 2^{|\mathcal{D}|}$ is the number of subsessions.

We note that these conditions hold for a large class of utility functions, including $U_d(y_d) = y_d^{(1-\phi)}/(1 - \phi)$ for $\phi > 0$ that is known to capture a large class of fairness criteria (see [19]). We remark that the strictness of the concavity of $U_d(\cdot)$ can be relaxed without too much complication. Also, observe that although $U_d(\cdot)$ is a strictly concave function of y_d, it is only concave in $\mathbf{x}$ since $\mathbf{y}$ and $\mathbf{x}$ are related as in (1).

Our goal is to design a practical algorithm that finds the optimal subsession rates $\hat{\mathbf{x}} := (\hat{x}_k)_{k=1}^{K}$ that maximize the aggregate utilization of the destinations. In order to formulate this problem more rigorously, we let $f_{ij}^{(d, \mathcal{D}_k)}$ be the *information flow rate* over link $(i, j) \in \mathcal{L}$ for an individual destination $d \in \mathcal{D}_k$ in Subsession k; and $r_{ij}^{\mathcal{D}_k}$ be the *physical flow rate* for Subsession k over link $(i, j) \in \mathcal{L}$. Under the intra-session network coding capability for each subsession, the physical and information flow rates are related by ([12])

$$r_{ij}^{\mathcal{D}_k} \ge \max_{d \in \mathcal{D}_k} f_{ij}^{(d, \mathcal{D}_k)}, \text{ for all } (i, j) \in \mathcal{L}, k \in \{1, \cdots, K\}. \quad (2)$$

Network coding enables us to transmit the coded packets at a rate equal to the maximum information rate of all the destination nodes belonging to that subsession. Information flow rate out of node i for Subsession k and $d \in \mathcal{D}_k$, is given by

$$f_{out(i)}^{(d, \mathcal{D}_k)} := \sum_{j:(i,j) \in \mathcal{L}} f_{ij}^{(d, \mathcal{D}_k)}, \quad \forall d \in \mathcal{D}_k, i \in \mathcal{N}.$$

Similarly, information flow rate into node i for subsession k

and $d \in \mathcal{D}_k$, is given by

$$f_{in(i)}^{(d,\mathcal{D}_k)} := \sum_{j:(j,i)\in\mathcal{L}} f_{ij}^{(d,\mathcal{D}_k)}, \quad \forall d \in \mathcal{D}_k, i \in \mathcal{N}.$$

Finally, we define the variable

$$x_i^k := \begin{cases} x_k, & \text{if } i = s, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Now, we are ready to formulate our goal as an optimization problem. We approach the problem from the destinations' point of view and maximize the destination utilities rather that subsession utilities while retaining the multi-rate multicast nature of information flow. We note that this is the natural formulation since the utility functions are described per destination. This leads the following optimization problem.

$$\max_{\{\mathbf{x},\mathbf{f},\mathbf{r}\geq\mathbf{0}\}} \quad \sum_{d\in\mathcal{D}} U_d(y_d(\mathbf{x})) \tag{4}$$

$$s.t. \quad f_{out(i)}^{(d,\mathcal{D}_k)} - f_{in(i)}^{(d,\mathcal{D}_k)} = x_i^k, \qquad i \neq d, \forall k, \forall d \in \mathcal{D}_k,$$

$$f_{ij}^{(d,\mathcal{D}_k)} \leq r_{ij}^{\mathcal{D}_k}, \qquad \forall k \in \{1,\cdots,K\}, \forall d \in \mathcal{D}_k,$$

$$\sum_{k=1}^{K} r_{ij}^{\mathcal{D}_k} \leq c_{ij}, \qquad \forall (i,j) \in \mathcal{L}.$$

The first constraint is the information flow conservation constraint per node, implying that the information influx to a node must be equal the information outflux, unless the node is a destination node. The second constraint is a network coding constraint indicating the relationship between information flow rate and physical flow rate given by equation (2). And, the final constraint is the link capacity constraint.

Under the stated assumptions on the utility function this optimization is a convex optimization problem, and strong duality holds due to the polyhedral nature of the constraint set (see [2]). However, since the utility function is not strictly concave in $\mathbf{x}$, there may not be a unique solution to the optimization problem (4). We denote the solution to the optimization problem by a set of subsession rates denoted by $\hat{\chi}$. Each point $\hat{\mathbf{x}} \in \hat{\chi}$ in the optimal set is a maximizer of the expression $\sum_{d\in D} U_d(y_d(\mathbf{x}))$.

In the following section, we use duality theory and dual methods to develop a flow control and resource allocation mechanism that solves the above problem. Then, in Section 5 we exploit the special structure of the problem to propose a nested multicasting strategy that significantly reduces the complexity of the original algorithm from exponential to linear in the number of destinations.

3. ALGORITHM DESIGN
In this section, we develop a joint congestion control, resource allocation and coding algorithm for the multi-rate multicast optimization problem (4). The development is based on a primal-dual update policy that guarantees optimal routes, stability, and optimal rate allocation.

Let $\mu := \left(\mu_i^{(d,\mathcal{D}_k)}\right)_{i,d,k}$ be the Lagrange multiplier vector associated with the first constraint of (4). Then, the La-

grangian is given by

$$L(\mathbf{x},\mu) = \sum_{d\in D} U_d(y_d)$$
$$- \sum_i \sum_k \sum_{d\in\mathcal{D}_k} \mu_i^{(d,\mathcal{D}_k)}(x_i^k - f_{out(i)}^{(d,\mathcal{D}_k)} + f_{in(i)}^{(d,\mathcal{D}_k)}),$$

where we dropped the $\mathbf{x}$ dependence of $y_d(\mathbf{x})$ for brevity, and $\mu_i^{(d,\mathcal{D}_k)}$ can be interpreted as *congestion price* at node i for Subsession k and Destination $d \in \mathcal{D}_k$. The congestion prices at different nodes can be indicative of queue lengths of the *logical queues* at each node i for subsession k and destination $d \in \mathcal{D}_k$. The design of physical queues from the logical queues is an issue of implementation which we do not discuss in this paper.

The Lagrangian can be re-arranged as

$$L(\mathbf{x},\mu) = \sum_{d\in\mathcal{D}} \left(U_d(y_d) - \sum_{k\ni d} \mu_s^{(d,\mathcal{D}_k)} x_k \right)$$
$$+ \sum_{(i,j)} \sum_k \sum_{d\in\mathcal{D}_k} f_{ij}^{(d,\mathcal{D}_k)}(\mu_i^{(d,\mathcal{D}_k)} - \mu_j^{(d,\mathcal{D}_k)}),$$

where s is the source node of the multi-rate multicast session. Then, the objective function of the dual problem becomes

$$\max_{\{\mathbf{x},\mathbf{f},\mathbf{r}\geq\mathbf{0}\}} L(\mathbf{x},\mu) = \max_{x\geq 0} \sum_{d\in\mathcal{D}} \left(U_d(y_d) - \sum_{k\ni d} \mu_s^{(d,\mathcal{D}_k)} x_k \right) \tag{5}$$

$$+ \max_{\left\{ \substack{0\leq f_{ij}^{(d,\mathcal{D}_k)} \leq r_{ij}^{\mathcal{D}_k} \\ \sum_{k\in K} r_{ij}^{\mathcal{D}_k} \leq c_{ij}} \right\}} \sum_{(i,j)} \sum_k \sum_{d\in\mathcal{D}_k} f_{ij}^{(d,\mathcal{D}_k)}(\mu_i^{(d,\mathcal{D}_k)} - \mu_j^{(d,\mathcal{D}_k)}).$$

$$\tag{6}$$

The above formulation enables us to break down the problem into two separate convex optimization problems: the first maximization (5) over $\mathbf{x}$ is related to the rate allocation; and the second maximization (6) over $(\mathbf{f},\mathbf{r})$ is related to scheduling, routing, and coding.

It should be noted that contrary to many earlier works in the literature (e.g. [14, 18, 16, 8, 23, 20, 5]) the nature of the rate allocation problem does not readily lend itself to complete decomposition across subsession rates. In this case, due to the network coding capabilities and the destination-based utility functions, the optimal subsession rates are interdependent. This motivates us to use first-order optimization methods to solve for the rate allocation (primal) problem, and iteratively update the rates in the subgradient direction.

On the other hand, the nature of the resource allocation problem (6) is similar to the backpressure schemes that are extensively studied in the literature (e.g. [24, 21, 10, 9, 17]), except that the objective function is defined in terms of "information flow rates" rather than the traditional physical link rates due to the network coding capability within subsessions. Thus, the implementation of the algorithm requires the description of network coding operations to be performed at the nodes.

The following iterative cross-layer mechanism describes the details of our proposed algorithm.

DEFINITION 1 (CROSS-LAYER MECHANISM). *Assume that the algorithm operates in slotted time, and at a time slot t, $(\mathbf{x}[t], \mu[t])$ is known.*

Rate Control: *The flow rate of each subsession k is updated depending upon the congestion level it observes through the price level at its source as follows: for all k,*

$$x_k[t+1]$$
$$= \left(x_k[t] + \alpha \sum_{d \in \mathcal{D}_k} \left(U_d'(y_d[t]) - \mu_s^{(d,\mathcal{D}_k)}[t] \right) \right)_m^M, \quad (7)$$

where $U_d'(y) = \frac{dU_d(y)}{dy}$; $y_d[t] := \sum_{k \ni d} x_k[t]$; and z_a^b denotes a projection of z to the closest point in the interval $[a,b]$. Further, $\alpha > 0$ is a small step-size parameter, $0 < m < \min_s \hat{x}_k$ where $\hat{\mathbf{x}}$ is the optimal solution, and M is a finite constant that is greater than $\sum_{(i,j) \in \mathcal{L}} c_{ij}$.

Resource Allocation and Coding: *For each link (i,j), two parameters $w_{i,j}^*[t]$ and $k_{i,j}^*[t]$ are computed as follows.*

$$w_{i,j}^*[t] = \max_k \sum_{d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)}[t] - \mu_j^{(d,\mathcal{D}_k)}[t])^+$$

$$k_{i,j}^*[t] = \arg\max_k \sum_{d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)}[t] - \mu_j^{(d,\mathcal{D}_k)}[t])^+$$

where $(z)^+ := \max(0, z)$. Here, $w_{i,j}^*[t]$ can be interpreted as the weight of the link (i,j), and $k_{i,j}^*[t]$ denotes the index of the subsession data to be transmitted over link (i,j). This means that the physical rate $r_{ij}^{\mathcal{D}_k}[t]$ of subsession $k_{i,j}^*[t]$ is equal to c_{ij}, and all the other subsessions are unscheduled.

The subsession $k_{i,j}^*[t]$ is served as follows:

1. *Let*
   $$\mathcal{D}_{k_{i,j}^*[t]}^+ = \{d \in \mathcal{D}_{k_{i,j}^*[t]} : \mu_i^{(d,\mathcal{D}_k)^*}[t] - \mu_j^{(d,\mathcal{D}_k)^*}[t] \geq 0\}.$$

   Then, for each $d \in \mathcal{D}_{k_{i,j}^*[t]}^+$ take $c_{i,j}$ packets from its associated queue. If the queue empties, continue to the next step.

2. *Form c_{ij} independent random linear combinations of these selected packets and transmit them over link (i,j).*

3. *At the receiver end, enqueue a copy of the incoming packets at each of the queues $\mu_j^{(d,\mathcal{D}_k)^*}$ with $d \in \mathcal{D}_{k_{i,j}^*}^+[t]$.*

Finally, update the μ vector as follows: for all i, k,

$$\mu_i^{(d,\mathcal{D}_k)}[t+1]$$
$$= \left(\mu_i^{(d,\mathcal{D}_k)}[t] + \beta(x_i^k[t] + f_{in(i)}^{(d,\mathcal{D}_k)}[t] - f_{out(i)}^{(d,\mathcal{D}_k)}[t]) \right)^+, \quad (8)$$

where $\beta > 0$ is a small stepsize parameter.

We remark that the dual update equation (8) in the cross-layer mechanism is tightly related to the actual queue-length

evolution. In particular, if we let $q_i^{(d,\mathcal{D}_k)}[t]$ denote the length of the queue at node i that contains packets destined to node $d \in \mathcal{D}_k$ as part of the subsession k, then, we approximately have $q[t] \approx \beta\mu[t]$. This means that a scaled version of the actual queue-lengths can also be used in the scheduling algorithm, and no extra price maintenance is necessary.

4. ALGORITHM ANALYSIS

In this section, we introduce a heuristic fluid model of the joint scheduler-congestion control mechanism of Definition 1, and prove its stability and optimality properties. The analysis of such a model sheds light into the operation of the original model without the complications associated with the discrete-time operation. It is shown in several earlier works that the optimality of such a fluid model suggests the optimality of the original discrete-time implementation (e.g. [8, 9, 17]). Here, we will only prove the optimality of the fluid model, leaving the analysis of the discrete-time implementation to an extended version of the paper.

To construct the fluid model we first approximate the operation of the discrete-time cross-layer algorithm by a heuristic continuous-time dynamic system. Accordingly, the evolution of the instantaneous subsession rates is assumed to be governed by the following differential equation: for all $k \in K$,

$$\dot{x}_k(t) = \alpha \left(\sum_{d \in \mathcal{D}_k} (U_d'(y_d(t)) - \mu_s^{(d,\mathcal{D}_k)}(t)) \right)_{x_k(t)}^+. \quad (9)$$

Similarly, the evolution of the prices is governed by the differential equation: for all i, k,

$$\dot{\mu}_i^{(d,\mathcal{D}_k)}(t) = \beta \left(x_i^k(t) + f_{in(i)}^{(d,\mathcal{D}_k)}(t) - f_{out(i)}^{(d,\mathcal{D}_k)}(t) \right)_{\mu_i^{(d,\mathcal{D}_k)}(t)}^+, \quad (10)$$

where $(y)_z^+$ is equal to 0 if $\{y = 0, z < 0\}$, and y, otherwise. Here, (t) is used instead of $[t]$ to signify that we are working in continuous-time. We assume that the resource allocation and coding algorithm computes the link schedules and rates at every instant of time as described in Section 3.

Notice that equations (10) and (9) mimic the evolution of (8) and (7), respectively. Thus, proving the optimality characteristics of this continuous-time system is important in extending its conclusion to the original system. The following theorem proves the globally asymptotic optimality of the continuous-time system.

THEOREM 1. *Starting from any initial condition $(\mathbf{x}(0), \mu(0))$, the state of the system $(\mathbf{x}(t), \mu(t))$ converges to $(\hat{\mathbf{x}}, \hat{\mu})$ as $t \to \infty$, where $\hat{x}_k$ is an element of the vector $\hat{\mathbf{x}} \in \hat{\chi}$ and $\hat{\mu}$ is an element in the set of optimal prices.*

PROOF. Consider the Lyapunov function

$$W_t(\mathbf{x}, \mu) = \frac{1}{2\alpha} \sum_{k \in K} (x_k(t) - \hat{x}_k)^2 +$$
$$\frac{1}{2\beta} \sum_i \sum_k \sum_{d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)}(t) - \hat{\mu}_i^{(d,\mathcal{D}_k)})^2$$

Differentiating the Lyapunov function with respect to time, we get the Lyapunov drift

$$\dot{W}_t(\mathbf{x}, \mu) = \frac{1}{\alpha} \sum_{k \in K} (x_k(t) - \hat{x}_k)\dot{x}(t)$$
$$+ \frac{1}{\beta} \sum_{i,k,d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)}(t) - \hat{\mu}_i^{(d,\mathcal{D}_k)})\dot{\mu}_i^{(d,\mathcal{D}_k)}(t)$$

For notational simplicity, we drop the time index but all the rate and flow quantities must be interpreted as time dependent functions. Using equation (9) and (10) to determine the primal and dual gradients, we get

$$\dot{W}_t(\mathbf{x}, \mu) = \sum_{k \in K} (x_k - \hat{x}_k)\Big(\sum_{d \in \mathcal{D}_k} (U_d^{'}(y_d) - \mu_s^{(d,\mathcal{D}_k)})\Big)^+$$
$$+ \sum_{i,k,d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)} - \hat{\mu}_i^{(d,\mathcal{D}_k)})\Big(x_i^k + f_{in(i)}^{(d,\mathcal{D}_k)} - f_{out(i)}^{(d,\mathcal{D}_k)}\Big)^+$$
$$\leq \sum_{k \in K} (x_k - \hat{x}_k)\Big(\sum_{d \in \mathcal{D}_k} (U_d^{'}(y_d) - \mu_s^{(d,\mathcal{D}_k)})\Big)$$
$$+ \sum_{i,k,d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)} - \hat{\mu}_i^{(d,\mathcal{D}_k)})\Big(x_i^k + f_{in(i)}^{(d,\mathcal{D}_k)} - f_{out(i)}^{(d,\mathcal{D}_k)}\Big)$$

The inequality follows since we remove the projection operation from the rate and price gradients.

Optimality of $(\hat{\mathbf{x}}, \hat{\mu})$ implies that

$$\sum_{d \in \mathcal{D}_k} (U_d^{'}(\hat{y}_d) - \hat{\mu}_s^{(d,\mathcal{D}_k)}) = 0, \qquad \forall k \in K.$$

We add this quantity in the first summation and optimal rate in the second summation of the Lyapunov drift. Hence,

$$\dot{W}_t(\mathbf{x}, \mu) = \sum_{k \in K} (x_k - \hat{x}_k)\Big(\sum_{d \in \mathcal{D}_k} [U_d^{'}(y_d) - U_d^{'}(\hat{y}_d)$$
$$+ \hat{\mu}_s^{(d,\mathcal{D}_k)} - \mu_s^{(d,\mathcal{D}_k)}]\Big) \quad (11)$$
$$+ \sum_{i,k,d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)} - \hat{\mu}_i^{(d,\mathcal{D}_k)})(x_i^k + f_{in(i)}^{(d,\mathcal{D}_k)} - f_{out(i)}^{(d,\mathcal{D}_k)}$$
$$+ \hat{x}_i^k - \hat{x}_i^k) \quad (12)$$

Consider the quantity,

$$\sum_i \sum_k \sum_{d \in \mathcal{D}_k} (\mu_i^{(d,\mathcal{D}_k)} - \hat{\mu}_i^{(d,\mathcal{D}_k)})(x_i^k - \hat{x}_i^k)$$
$$= \sum_k \sum_{d \in \mathcal{D}_k} (\mu_s^{(d,\mathcal{D}_k)} - \hat{\mu}_s^{(d,\mathcal{D}_k)})(x_k - \hat{x}_k).$$

The equality follows from the definition of x_i^k in (3). This quantity cancels with the second part of (11) and hence the Lyapunov drift expression reduces to

$$\dot{W}(\mathbf{x}, \mu) = \sum_{k \in K} (x_k - \hat{x}_k)\Big(\sum_{d \in \mathcal{D}_k} (U_d^{'}(y_d) - U_d^{'}(\hat{y}_d))\Big) \quad (13)$$
$$+ \sum_{i,k,d \in \mathcal{D}_k} \hat{\mu}_i^{(d,\mathcal{D}_k)}(f_{out(i)}^{(d,\mathcal{D}_k)} - f_{in(i)}^{(d,\mathcal{D}_k)} - \hat{x}_i^k) \quad (14)$$
$$+ \sum_{i,k,d \in \mathcal{D}_k} \mu_i^{(d,\mathcal{D}_k)}(\hat{x}_i^k + f_{in(i)}^{(d,\mathcal{D}_k)} - f_{out(i)}^{(d,\mathcal{D}_k)}) \quad (15)$$

Next, note that (13) can be rearranged as

$$(13) = \sum_{d \in D} \Big((y_d - \hat{y}_d)(U_d^{'}(y_d) - U_d^{'}(\hat{y}_d))\Big). \quad (16)$$

Observe that $(16) \leq 0$ since the utility function is concave by assumption.

To show that other terms in the Lyapunov drift are negative, we start with (14). We can write

$$\sum_i \sum_k \sum_{d \in \mathcal{D}_k} (\hat{\mu}_i^{(d,\mathcal{D}_k)})(f_{out(i)}^{(d,\mathcal{D}_k)} - f_{in(i)}^{(d,\mathcal{D}_k)})$$
$$= \sum_{(i,j)} \sum_k \sum_{d \in \mathcal{D}_k} f_{ij}^{(d,\mathcal{D}_k)}(\hat{\mu}_i^{(d,\mathcal{D}_k)} - \hat{\mu}_j^{(d,\mathcal{D}_k)})$$
$$\leq^{(a)} \sum_{k \in K} \sum_{d \in \mathcal{D}_k} \hat{x}_k\hat{\mu}_s^{(d,\mathcal{D}_k)},$$

where the inequality (a) can be proved by the following argument: by Karush-Kuhn-Tucker (KKT) conditions (see [2])

$$\hat{\mu}_i^{(d,\mathcal{D}_k)}(\hat{x}_i^k + \hat{f}_{in(i)}^{(d,\mathcal{D}_k)} - \hat{f}_{out(i)}^{(d,\mathcal{D}_k)}) = 0,$$
$$\forall i \in N, k \in K$$

Summing over all nodes and subsessions and noting that $\hat{x}_i^k = 0, \quad \forall i \neq s$,

$$\sum_{k \in K} \sum_{d \in \mathcal{D}_k} \hat{\mu}_s^{(d,\mathcal{D}_k)}\hat{x}_k$$
$$= \sum_i \sum_k \sum_{d \in \mathcal{D}_k} \hat{\mu}_i^{(d,\mathcal{D}_k)}(\hat{f}_{out(i)}^{(d,\mathcal{D}_k)} - \hat{f}_{in(i)}^{(d,\mathcal{D}_k)})$$
$$= \sum_{(i,j)} \sum_k \sum_{d \in \mathcal{D}_k} \hat{f}_{ij}^{(d,\mathcal{D}_k)}(\hat{\mu}_i^{(d,\mathcal{D}_k)} - \hat{\mu}_j^{(d,\mathcal{D}_k)})$$
$$\geq \sum_{(i,j)} \sum_k \sum_{d \in \mathcal{D}_k} f_{ij}^{(d,\mathcal{D}_k)}(\hat{\mu}_i^{(d,\mathcal{D}_k)} - \hat{\mu}_j^{(d,\mathcal{D}_k)})$$

for any feasible f satisfying $0 \leq f_{ij}^{(d,\mathcal{D}_k)} \leq r_{ij}^{\mathcal{D}_k}$ and $\sum_{k \in K} r_{ij}^{\mathcal{D}_k} \leq c_{ij} \ \forall (i,j)$, since $(\hat{\mathbf{x}}, \hat{f}, \hat{\mu})$ maximizes the Lagrangian function in (6). From the result of (17), it follows that $(14) \leq 0$.

To prove $(15) \leq 0$, we proceed as follows: from the flow balance equations we have

$$\hat{x}_i^k = \hat{f}_{out(i)}^{(d,\mathcal{D}_k)} - \hat{f}_{in(i)}^{(d,\mathcal{D}_k)}, \quad \forall i \in N, k \in K, d \in \mathcal{D}_k.$$

Multiplying by corresponding subsession prices and summing it over all nodes and subsessions

$$\sum_{k \in K} \sum_{d \in \mathcal{D}_k} \hat{x}_k\mu_s^{(d,\mathcal{D}_k)} = \sum_i \sum_k \sum_{d \in \mathcal{D}_k} (\hat{f}_{out(i)}^{(d,\mathcal{D}_k)}$$
$$- \hat{f}_{in(i)}^{(d,\mathcal{D}_k)})\mu_i^{(d,\mathcal{D}_k)}$$
$$= \sum_{(i,j)} \sum_k \sum_{d \in \mathcal{D}_k} \hat{f}_{ij}^{(d,\mathcal{D}_k)}(\mu_i^{(d,\mathcal{D}_k)} - \mu_j^{(d,\mathcal{D}_k)})$$
$$\leq \sum_{(i,j)} \sum_k \sum_{d \in \mathcal{D}_k} f_{ij}^{(d,\mathcal{D}_k)}(\mu_i^{(d,\mathcal{D}_k)} - \mu_j^{(d,\mathcal{D}_k)}) \quad (17)$$

Inequality follows since the resource allocation operation in Definition 1 sets the information flow rates to maximize the expression in (17).

From (17) it follows that $(15) \leq 0$. Combining the results, $(13) + (14) + (15) \leq 0$, and hence the Lyapunov drift,

$\dot{W}_t(\mathbf{x}, \mu) \leq 0$ and further,

$$\xi := \{(\mathbf{x}, \mu) : \dot{W}_t(\mathbf{x}, \mu) = 0\}$$

is contained in the set

$$S := \{(\mathbf{x}, \mu) : (13) = (14) = (15) = 0\}$$

Let M be the largest invariant set of the primal-dual algorithm contained in ξ. By LaSalle's invariance principle (provided in the Appendix for completeness), $(\mathbf{x}(t), \mu(t))$ converges to the set M as $t \rightarrow \infty$. Since $M \subset \xi \subset S$, as $t \rightarrow \infty$, the limit point of the pair $(\mathbf{x}(t), \mu(t))$ must also satisfy $(13) = (16) = 0$. It must be noted that strict concavity of the utility functions implies $(16) = 0$ only when $U_d^{'}(y_d) = U_d^{'}(\hat{y}_d)$ and hence $\lim_{t \rightarrow \infty} y_d(t) = \hat{y}_d$. $\quad\square$

The arguments in the proof of this theorem can be used to form the foundation of the discrete-time analysis, which is omitted in this paper due to space constraints.

While such optimality characteristics is important and certainly attractive, it must be noted that our cross-layer algorithm is based on decomposing a single multi-rate session with $|\mathcal{D}|$ number of destinations into $2^{|\mathcal{D}|}$ single-rate multicast subsessions, one for each subset of destinations. Thus, the complexity of the algorithm grows exponentially with the size of the multicast session. Next, we will propose a low-complexity implementation by utilizing the special structure of the problem.

5. COMPLEXITY REDUCTION

In this section, we propose a novel strategy to reduce the exponential complexity of the cross-layer mechanism with respect to the number of destination nodes. The main idea behind our approach can be described by a horizontal water-filling analogy as follows. Consider the maximum achievable rates in the multi-rate multicast session as a function of the destinations, such illustrated in Figure 1, where we can see that Destination 4 has the best channel, while Destination 5 has the worst channel.
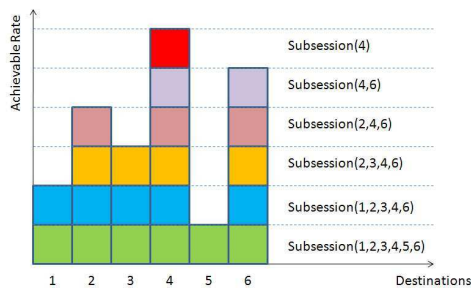
Figure 1: Determining the optimal subsessions with a nested structure.

The idea is to start by serving a single subsession containing all the destinations and then adding more subsessions by gradually eliminating the bottleneck destinations by utilizing the pricing information at the sources. In the example of Figure 1, we see that Destination 5 must be eliminated in the first round as it is the bottleneck link of the first (i.e. bottom) subsession. Then, in the second round the

first two subsessions are served, leading to the identification of Destination 1 as the bottleneck destination in the second subsession. Thus, Destination 1 is eliminated in the third round to construct the third subsession, and so on. The details of our proposed procedure is described next.

DEFINITION 2 (LOW-COMPLEXITY MECHANISM). *Perform the following steps.*

• *Initialize the number of subsessions R to 1, with $\mathcal{D}_1 = \mathcal{D}$.*

• *In the Rth round, $R = 1, \cdots, K - 1$, with $\{\mathcal{D}_1, \cdots, \mathcal{D}_R\}$ denoting the existing multicast subsessions, do:*

(i) Run the cross-layer mechanism of Definition 1 only for the subsessions with destinations $\{\mathcal{D}_1, \cdots, \mathcal{D}_R\}$, such that $x_r(t) \rightarrow \tilde{x}_r$, for each $r = 1, \cdots, R$; and $\mu_s^{(d, \mathcal{D}_r)}(t) \rightarrow \tilde{\mu}_s^{(d, \mathcal{D}_r)}$, for each $r = 1, \cdots, R$, and $d \in \mathcal{D}_r$.

(ii) Let $d_R^\star = \arg\max_{d \in \mathcal{D}_R} \tilde{\mu}_s^{(d, \mathcal{D}_R)}$, which implies that $d_R^\star$ is the bottleneck destination in the multicast subsession $\mathcal{D}_R$ since its price is the maximum.

(iii) Construct the $(R+1)^{st}$ subsession as $\mathcal{D}_{R+1} = \mathcal{D}_R \backslash d_R^\star$. If $(R+1) = K$, exit. Otherwise start the $(R+1)^{st}$ round.

It can be seen in Step (ii) of the iterative procedure that the limit price levels at the source are used to identify the bottleneck link in the most recent, i.e. R^{th}, multicast subsession, and in Step (iii), the next subsession is created by eliminating the identified bottleneck, $d_R^\star$. The exact analysis of this scheme is complicated and beyond the scope of this work. Instead, in the next section, we simulate several network settings to compare the optimality characteristics of the original scheme and the low-complexity variation. Our simulations demonstrate that the low complexity scheme continues to achieve full optimality in addition to the complexity gains it provides.

6. NUMERICAL RESULTS

In this section we provide some simulation results for three well-known network topologies serving a multi-rate multicast session with 2, 3, and 6 destinations, respectively. We study the performance of the original cross-layer algorithm of Definition 1 and the low-complexity implementation proposed in Definition 2 under various capacity constraints. We start with the canonical butterfly network.

6.1 Butterfly Network

Consider the butterfly network as shown in Figure 2. Assume that there is a single multi-rate multicast session having node s as the source node and nodes d_1 and d_2 as the destination nodes. We decompose this multi-rate multicast session into three single-rate multicast subsessions each having source node s. The subsessions have the following destination set, $\mathcal{D}_1 = \{d_1\}$, $\mathcal{D}_2 = \{d_2\}$ and $\mathcal{D}_3 = \{d_1, d_2\}$. We interpret subsessions 1 and 2 to be unicast subsessions and 3 to be the multicast subsession. We assume that the utilization achieved at each destination $d \in \mathcal{D}$, is measured by $U_d(y_d) = \log(y_d)$, which leads to the proportionally fair

allocation. We choose the α and β parameters to be 0.01, and examine the network under different scenarios to show the effectiveness of our cross-layer algorithm.
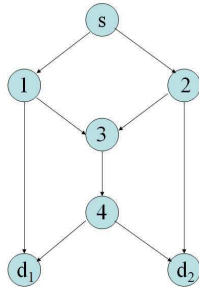

Figure 2: The butterfly network.

Case(a): Initially we analyze the network with link capacity of all links to be 1 unit. The subsession rates at convergence point are tabulated in Table 1 and the time evolution of the rates is plotted in Figure 3. The rows of the table indicate different subsessions and their corresponding destination set. "ON" indicates the node is in the destination set and "OFF" indicates it is not. We see that the rate values at convergence point are 0, 0 and 2 respectively. Each destination receives a data rate of 2 which equals the theoretically maximum achievable rate specified by the max-flow-min-cut theorem. Also, it can be observed that rates for the unicast subsessions are 0. Hence when the link capacities are uniform, all the data can be transmitted through multicast taking full advantage of network coding.

Case(b): In the second simulation scenario, we introduce a bottleneck link in the paths of one of the destinations, i.e. we reduce the average capacity over link $(1, d_1)$ to 0.1 units. We achieve this average capacity through a Bernoulli random process with parameter 0.1. The subsession rates are again tabulated in Table 2 and evolution of rates in Figure 4. The net achieved rate for destination d_1 is 1.1 units where as destination d_2 receives a rate of 2 units. Case (b) clearly shows the effectiveness of our algorithm: while single-rate multicast with fixed coding subgraphs would have resulted in a lower data rate to both the destinations due to receiver d_1 creating the bottleneck for the whole multicast group, our cross layer mechanism still achieves rates very close to theoretical value for either destinations.

Table 1: Case (a)

	d_1	d_2	x_k
S 1	OFF	ON	0
S 2	ON	OFF	0
S 3	ON	ON	2

Table 2: Case (b)

	d_1	d_2	x_k
S 1	OFF	ON	0
S 2	ON	OFF	0.9
S 3	ON	ON	1.1

Case(c): In the third simulation scenario, we introduce bottleneck links in the paths of both the destinations by setting the average capacities of links $(1, d_1)$ and $(2, d_2)$ to 0.1. The resulting rate values at convergence are tabulated in Table 3 and evolution of rates in Figure 5. It can be seen once again that all the data is transmitted through the multicast subsession and the algorithm takes advantage of network coding

Table 3: Case (c)

	d_1	d_2	x_k
S 1	OFF	ON	0
S 2	ON	OFF	0
S 3	ON	ON	1

Table 4: Case (d)

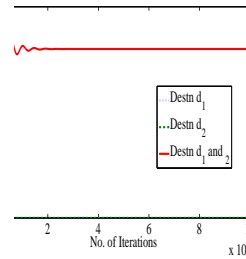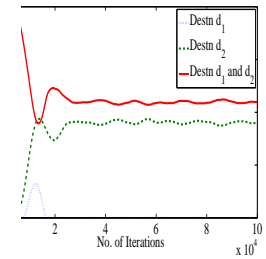|	d_1	d_2	x_k
S 1	OFF	ON	0.5
S 2	ON	OFF	0.5
S 3	ON	ON	0.6

Figure 3: Case (a)
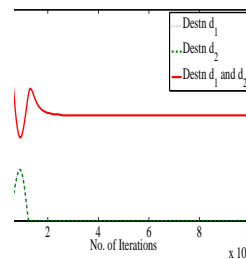

Figure 4: Case (b)
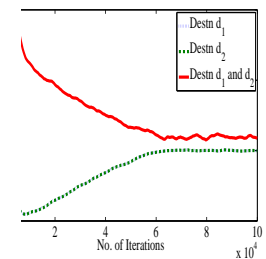

Figure 5: Case (c)

Figure 6: Case (d)

between packets of the multicast subsession to achieve a rate of 1 unit to either of the destinations.

Case(d): In the fourth simulation scenario, we set the average capacity of link $(3, 4)$ to 0.1 units, and all others to 1. The resulting rate values at convergence are tabulated in Table 4 and evolution of rates in Figure 6. Both the destinations achieve a rate of 1.1 units, which is optimal. We note that the proposed rates of 0.5 units for unicast and 0.6 units for multicast subsessions can be achieved by time sharing the links $(1, d_1)$ and $(2, d_2)$ between the unicast and multicast subsessions.

Also, note that the system converges to within a small neighborhood of the optimal, since we have chosen constant stepsizes. The choice of the step-size parameter determines the tradeoff between optimality and convergence rate. In particular, the smaller the stepsize, the slower the convergence and the closer to the optimal, which is a general characteristic of any gradient based method.

6.2 Larger Networks

In this section, we illustrate the effectiveness of the low complexity multicasting strategy of Definition 2 not only in reducing the algorithm complexity but also in achieving data rates very close to the optimal value.

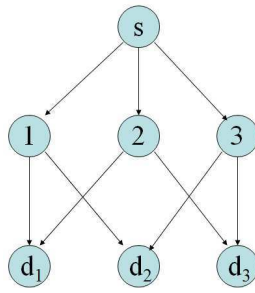The network considered for illustrating complexity reduc-

Figure 7: Network serving a session with 3 destinations.

Figure 9: Network serving a session with 6 destinations.

tion is shown in Figure 7. The multi-rate session has three destinations, and all the links are assumed to have unit capacity, except for link $(1, d_1)$, which has an average capacity of 0.1 units. The rate values at convergence for different subsessions at different rounds are tabulated in the rightmost column of Table 5 and the evolution of rates for the last round is plotted in Figure 8. The numbers below the destinations in Table 5 indicate the convergent price values at the source node for the corresponding destination. The boxed numbers in each round indicate that the corresponding destination has the highest price and is the bottleneck receiver in the subsession. Hence it is eliminated in the subsequent rounds.

Table 5: Subsession rates at different rounds for the nested multicast strategy.

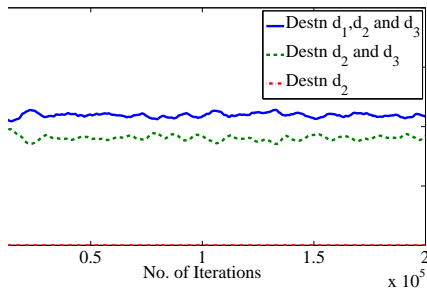| Round #	d_1	d_2	d_3	x_k
Round 1	2.7	0.0	0.0	1.1
Round 2	0.9	0.6	0.4	1.1
	-	0.5	0.5	0.9
Round 3	1.0	0.7	0.3	1.0
	-	0.5	0.5	0.9
	-	0.5	-	0

Figure 8: Evolution of rates in the third and final round of the Network in Figure 7.

Next, we consider an even larger network serving a session with six destination nodes as shown in Figure 9. The link capacities for all the links are 1 unit except for those indicated in Figure 9. The prices and rate values at convergence for different subsessions at different rounds are tabulated in Table 6 and the evolution of rates for the last round is
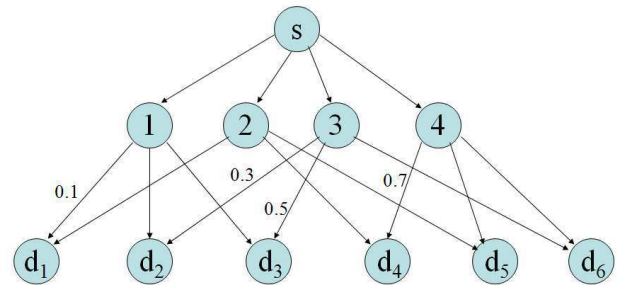
plotted in Figure 10. The optimality and effectiveness of the algorithm is clearly evident in this case: the original cross layer mechanism has to deal with an exponential number of subsessions (64 in this case) and becomes infeasible as the number of destinations increases, whereas the low-complexity implementation only uses up to a linear number of subsessions (6 in this case).

Table 6: Subsession rates at different rounds for the nested multicast strategy.

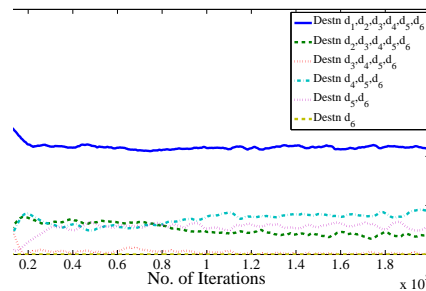| Round#	d_1	d_2	d_3	d_4	d_5	d_6	x_k
Round 1	5.2	0.0	0.0	0.0	0.0	0.0	1.2
Round 2	0.9	3.4	0.1	0.0	0.0	0.0	1.1
	-	3.3	0.1	0.0	0.0	0.0	0.4
$\vdots$	$\vdots$				$\vdots$	$\vdots$
$\vdots$	$\vdots$				$\vdots$	$\vdots$
Round 6	0.9	1.6	0.0	0.6	0.0	0.9	1.1
	-	0.9	0.7	0.7	0.7	0.2	0.2
	-	-	0.8	0.7	0.7	0.1	0.0
	-	-	-	0.6	0.5	0.5	0.5
	-	-	-	-	0.5	0.5	0.3
	-	-	-	-	-	0.6	0

Figure 10: Evolution of rates in the third and final round of the Network in Figure 9.

7. CONCLUSIONS

In this paper, we formulated and solved the problem of multi-rate multicasting for aggregate utility maximization with network coding capabilities. Our approach is based on decomposing the multi-rate session into many single-rate sessions, each using network coding capabilities, and

then optimizing over the subsession rates. Using primal-dual methods, we developed a cross-layer algorithm which specifies the rate allocation, routing, scheduling, and coding decisions based on appropriately maintained dual parameters (prices), and then proved that a continuous-time model of our algorithm converges to the optimal allocation from any initial condition. Noting the exponential complexity of the original design with respect to the size of the multicast group, we utilized the special structure of problem to develop a low-complexity implementation that reduces the complexity to linear in the size of the multicast group without sacrificing significantly from optimality. This algorithm gradually constructs the set of subsessions with a nested destination set structure, instead of considering all of them at once. We also performed extensive simulations to confirm the optimality characteristics of the original algorithm and its low-complexity variant.

8. REFERENCES

[1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46:1204–1216, July 2000.

[2] D. Bertsekas, A. Nedich, and A. Ozdaglar. *Convex Analyis and Optimization*. Athena Scientific, Belmont, MA, 2003.

[3] L. Bui, R. Srikant, and A. Stolyar. Optimal resource allocation for multicast subsessions in multihop wireless networks. *Philosophical Transactions of the Royal Society*, 2008.

[4] L. Chen, T. Ho, S. H. Low, M. Chiang, and J. C. Doyle. Optimization-based rate control for multicast network coding. In *Proceedings of IEEE Infocom*, pages 1163–1171, Anchorage, AL, April 2007.

[5] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle. Jointly optimal congestion control, routing, and scheduling for wireless ad hoc networks. In *Proceedings of IEEE Infocom*, Barcelona, Spain, April 2006.

[6] S. Deb, M. Effros, T. Ho, D. R. Karger, R. Koetter, D. S. Lun, M. Médard, and N. Ratnakar. Network coding for wireless applications: A brief tutorial. In *Proc. International Workshop on Wireless Ad-hoc Networks (IWWAN) 2005*, May 2005. Invited paper.

[7] A. Eryilmaz and D. S. Lun. Control for inter-session network coding. Technical Report 2722, MIT LIDS, Aug. 2006.

[8] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length based scheduling and congestion control. In *Proceedings of IEEE Infocom*, volume 3, pages 1794–1803, Miami, FL, March 2005.

[9] A. Eryilmaz and R. Srikant. Resource allocation of multi-hop wireless networks. In *Proceedings of International Zurich Seminar on Communications*, February 2006.

[10] A. Eryilmaz, R. Srikant, and J. R. Perkins. Stable scheduling policies for fading wireless channels. *IEEE/ACM Transactions on Networking*, 13:411–425, April 2005.

[11] T. Ho, Y.-H. Chang, and K. Han. On constructive network coding for multiple unicasts. In *Proc. 44th Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2006.

[12] T. Ho and D. Lun. *Network Coding: An Introduction*. Cambridge University Press, 2008.

[13] K. Kar, S. Sarkar, and L. Tassiulas. A scalable low-overhead rate control algorithm for multirate multicast sessions. *IEEE Journal on Selected Areas in Communications*, 20(8):1541–1557, October 2002.

[14] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[15] R. Koetter and M. Médard. Beyond routing: An algebraic approach to network coding. *IEEE Transactions on Information Theory*, 11:782–795, October 2003.

[16] X. Lin and N. Shroff. Joint rate control and scheduling in multihop wireless networks. In *Proceedings of IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[17] X. Lin and N. Shroff. The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks. In *Proceedings of IEEE Infocom*, Miami, FL, March 2005.

[18] S. H. Low and D. E. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7:861–875, December 1999.

[19] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.

[20] M. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. In *Proceedings of IEEE Infocom*, pages 1723–1734, Miami, FL, March 2005.

[21] M. Neely, E. Modiano, and C. Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *Proceedings of IEEE Infocom*, pages 745–755, April 2003.

[22] S. Sarkar and L. Tassiulas. A framework for routing and congestion controlfor multicast information flows. *IEEE Transactions on Information Theory*, 48:2690–2708, 2002.

[23] A. Stolyar. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems*, 50(4):401–457, 2005.

[24] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 36:1936–1948, December 1992.

APPENDIX

THEOREM 2 (LASALLE'S INVARIANCE PRINCIPLE). *Consider the differential equation: $\dot{y}(t) = f(y(t))$. Let $Y : D \to R$ be a radially unbounded (i.e. $\lim_{||z||\to\infty} Y(z) = \infty$), continuously differentiable, positive definite function such that $\dot{Y}(z) \leq 0$ for all $z \in D$. Let ξ be the set of points in D where $\dot{Y}(z) = 0$. Let M be the largest invariant set[1] in ξ. Then, every solution starting in D approaches M as $t \to \infty$.*

---

[1] S is defined as an *invariant set* for a dynamic system $\dot{y}(t) = f(y(t))$ if every trajectory $y(t)$ which starts from a point in S always remains in S.