Dynamic Coding and Rate-Control for Serving Deadline-Constrained Traffic over Fading Channels

Harsha Gangammanavar and Atilla Eryilmaz

***Abstract*— We formulate and solve the problem of optimal dynamic coding and rate-control for broadcasting deadline-constrained traffic over time-varying wireless channels. In particular, we propose and analyze a novel policy that utilizes a combination of *pricing* and *dynamic programming* strategies to jointly optimize the operation of the following two components: (i) a dynamic rate allocation policy, which manages the incoming traffic flow rates so as to maximize their weighted sum; (ii) and a dynamic coding window selection policy, which satisfies the reliability metrics of deadline-constrained flows that are expressed in terms of delivery ratios. Our heuristic fluid analysis of the resulting stochastic network operation indicates that our joint policy maximizes the weighted sum of the flow deadline-constrained throughput subject to heterogeneous reliability requirements imposed on them. We also apply these general results to an important cellular downlink scenario with network coding capabilities to study its behavior under various conditions. Our simulations reveal that the dynamic coding strategy outperforms the optimal static coding strategy by opportunistically exploiting the statistical variations in the arrival and channel processes.**

***Index Terms*— Delay-aware dynamic coding, stochastic control, network coding, deadline-constrained throughput optimization.**

I. Introduction

While the traditional performance measure of a communication system is *throughput*, many real-world applications also have a range of delay-sensitivities and Quality-of-Service (QoS) requirements that are not accounted for. In particular, real-time media broadcasting or two-way voice/video communication applications possess very stringent deadline constraints and differing tolerance levels to dropped bits. Moreover, different flows entering the system may have different degrees of importance, necessitating prioritization of certain flows over the others.

Information theory reveals that there is a fundamental relationship between the reliable transmission rate and the *coding block* (also called the *coding window*) *size* used to map messages into transmission signals. In particular, the reliable transmission rate may be increased towards the capacity of the channel by increasing the coding window size. Increasing the coding window size also causes larger delay and in the presence of the aforementioned strictly deadline constrained traffic, becomes unacceptable beyond a level Thus, a radically different coding strategy must be employed by the transmitter to maximize the user satisfaction. Also, since the characteristics and requirements of the applications may change, the solution must be able to adapt to their changes.

In this work, we study the optimal service of such deadline-constrained flows in the presence of a various block coding strategies that the transmitter can employ. Each such strategy is described by a *completion time distribution*, that models the random time it takes for a coding window of a given size to be successfully transmitted over the communication channel. [1]. In this setup, we aim to find an optimal rate-control and coding strategy for serving deadline-constrained flows with varying reliability requirements (in term of delivery ratios) and priorities. To that end, our contributions can be summarized as follows:

- We provide a generic communication system model with a multi-timescale operation that accounts for the deadline-constraints of the incoming traffic. The system also includes the modeling of the heterogeneous reliability requirements (i.e., delivery ratios) and priorities of different deadline-constrained flows (Section II).
- We formulate the problem of coding and rate-control for maximizing a weighted-sum of the flow rates subject to both their short-term deadline constraints and their long-term reliability requirements (Section III). We note the difficulties in solving the resulting infinite-horizon stochastic control problem through traditional methods.
- We propose a novel dynamic algorithm that jointly determines the incoming flow rates and the coding operation in order to maximize the weighted sum of deadline-constrained flow rates of many flows with varying tolerance levels to packet drops (Section IV). This algorithm combines *dynamic pricing* at a slow timescale with *finite-horizon dynamic programming* at a fast timescale to achieve the formulated global objective. The global asymptotic optimality of a heuristic fluid model of this algorithm is provided in I.
- We analyze a continuous-time fluid approximation of the stochastic system operating under our proposed algorithm. We show that the system is *globally asymptotically optimal*, i.e., starting from any initial condition it evolves towards the optimal set of solutions of a deterministic problem that approximates the mean behavior of the original stochastic optimization problem (Appendix I).
- We apply the developed algorithm to an important application in cellular downlink scenario whereby a base station broadcasts multiple streaming deadline-constrained flows to N receivers over randomly varying erasure channels. We further study the performance of our dynamic algorithm with

H. Gangammanavar and A. Eryilmaz are with The Ohio State University, {{gangammh, eryilmaz}@ece.osu.edu}. This work was supported by DTRA grant HDTRA1-08-1-0016, and the Control-Based Mobile Ad-Hoc Networking (CBMANET) Program under DARPA subcontract no. 060786.

[1]We note that this is an alternative description of the coding performance to the traditional one that describes the probability of decoding error for a fixed transmission duration. We find that this alternative model is more useful for our design and analysis.

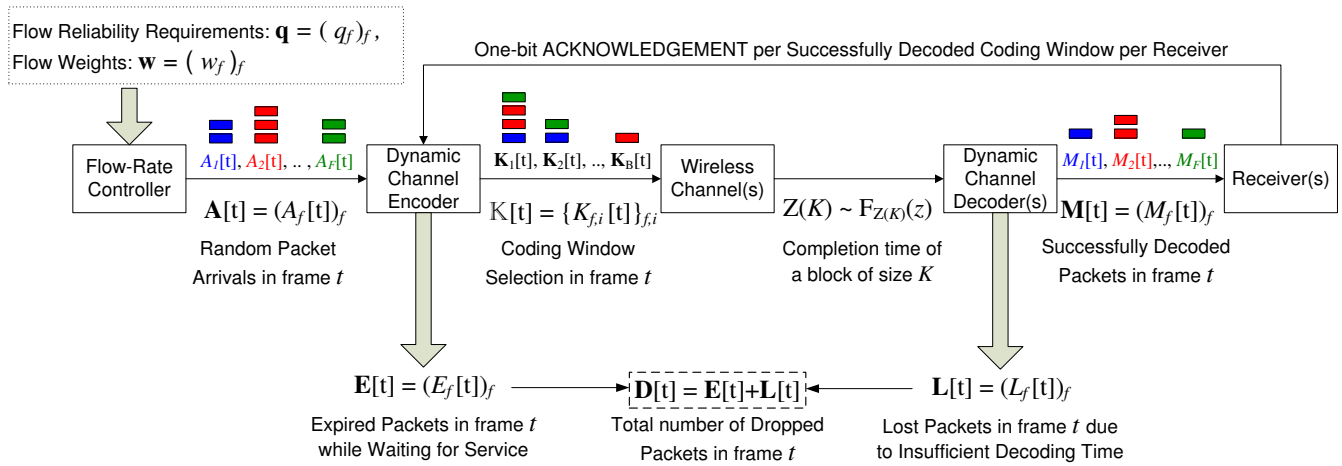

Fig. 1. Block diagram of the generic communication system for broadcasting deadline-constrained flows with varying reliability requirements and priorities.

and without *network coding* capabilities, and compare the dynamic policy performance to a static one to see strict improvements, even for small scenarios(Section V).

We note that queueing systems under impatient customers have been studied in the literature (e.g. [2], [10], [13]) for various cases of preemption, arrival/service rate distributions, etc. Yet, these works do not model the priorities and tolerance levels of applications, and do not account for possible coding parameters, and hence are not applicable to our problem. Also, recent works ([7], [8]) have studied the congestion-control and scheduling problem for similar deadline constrained traffic with reliability constraints. However, they also do not allow for coding flexibilities, which fundamentally changes the shape of the achievable rate region and calls for a dynamic strategy for optimizing over coding decisions that we develop.

II. System Model

We study the general communication system depicted in Figure 1, whereby a transmitter serves a set $\mathcal{F}$ of deadline-constrained flows over unreliable wireless channel(s). Next, we describe the system components and operation in detail. , and model the reliability and priority requirements of the flows. Then, we describe the main objective of this work in words, which is to be rigorously formulated and solved in the subsequent sections.

A. Description of the System Components

Flow-Rate Controller: The packets of each flow are generated according to the *Flow-Rate Controller* mechanism that is to be designed. Due to deadline constraints, the transmitted packets are useful at the receiver(s) only if they are successfully decoded within a constant τ time units after their generation. For simplicity, we assume that the system operates in fixed duration *time slots*, and hence from now on we will measure time in terms of time slots.

Multi-timescale Operation: We consider a multiple timescale operation of the system to allow the rate controller to operate at a slower time scale than the wireless channel variations (see Figure 2). Accordingly, we assume that the rate controller operates in a time scale of τ time slots, which we call a $frame$. In particular, at the beginning of Frame-t, the rate controller generates $A_f[t]$ number of Flow-f packets for that frame, which is allowed to be a random variable.

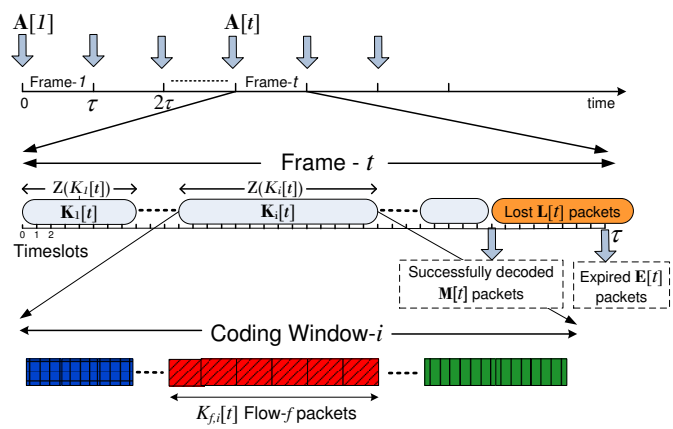

Fig. 2. Operation of the communication system of Figure 1 over time.

Dynamic Channel Encoder: The vector $\mathbf{A}[t] = (A_f[t])_f$ of total packet arrivals in Frame-t enter the *Dynamic Channel Encoder* that is allowed to perform block coding for reliable transmission. The encoder distributes the available packets into groups, called *coding windows*, for sequential transfer over the channel. We emphasize that the encoder only has access to the statistics of the channel(s), not their realizations.

We let the matrix $\mathbb{K}[t] = (K_{f,i}[t])_{f,i}$ denote the coding window selection in frame t, where $K_{f,i}[t]$ gives the number of Flow-f packets in the ith coding window. We also let $\mathbf{K}_i[t] = (K_{f,i}[t])_f$ denote the composition of the ith block in frame t. Furthermore, we let $K_i[t]$ and $K^f[t]$, respectively, denote the total number of packets in the ith block of frame t, and the total number of Flow-f packets scheduled for frame t.

Block Transmission and Decoding: Once the coding window for frame t, i.e. $\mathbb{K}[t]$, is selected, the constructed coding blocks $\{\mathbf{K}_1[t], \mathbf{K}_2[t], \cdots\}$ are sequentially transmitted over the wireless channel(s). In particular, the transmission of the second coding window starts only after the first coding window is successfully decoded by all the intended receivers. The successful decoding of each block is signaled through a single bit ACK signal by each receiver (see Figure 1).

We note that the inner system components, namely the channel encoder/decoder and the wireless channel, operate at

a time-scale of time slots (see Figure 2). Clearly, the amount of time required for the successful decoding of a coding block of size K, henceforth called the *completion time*, is a random variable that depends on the channel statistics and the coding strategy employed by the encoder. To capture this randomness, we let $Z(K)$ denote the completion time (in time slots) of a block of size K packets, which is generated according to a given distribution function $F_{Z(K)}(z)$. In Section V, we shall provide examples of such distribution functions in downlink broadcast setup over erasure channels.

Measure of Transmission Success or Failure: If the frame ends during the completion of a coding block transmission, we assume that all the packets in that block are lost. The number of lost packets in frame t is denoted by $\mathbf{L}[t] = (L_f[t])_f$. Similarly, the vector $\mathbf{M}[t] = (M_f[t])_f$ denotes the number of successfully decoded packets of each flow in frame t. Since the completion time is random, $\mathbf{M}[t]$ is also random and is described as

$$\begin{aligned} M_f[t] &= \sum_{i\geq 1} K_{f,i}[t] \cdot \mathcal{I}\left(\sum_{j=1}^{i} Z(K_j[t]) \leq \tau < \sum_{j=1}^{i+1} Z(K_j[t])\right) \\ &=: \quad g(\mathbf{Z}(\mathbb{K}[t])), \end{aligned} \tag{1}$$

where $\mathcal{I}(B)$ is the indicator function of event B. This expression reveals that $\mathbf{M}[t]$ is a complicated function $g(\mathbf{Z}(\mathbb{K}[t])$ of the completion times under the particular coding decision $\mathbb{K}[t]$. Thus, we denote the distribution of $\mathbf{M}[t]$ with $F_{g(\mathbf{Z}(\mathbb{K}[t]))}(m)$.

Note that some packets that are scheduled for coding and transmission may not even get a chance to start their transmission before the end of their frame. These packets are assumed to *expire* at the end of the frame, with $\mathbf{E}[t] = (E_f[t])_f$ denoting the number of such expiries. Hence, together with the losses, the total number of dropped packets $\mathbf{D}[t] = (D_f[t])$ in frame t equals $\mathbf{E}[t] + \mathbf{L}[t]$ (see Figure 1).

This completes the description of the communication system components of Figure 1. Next, we describe the preferences and requirements associated with the flows.

Application Preferences and Requirements: So far, we have only assumed that packet transmissions of each flow is considered successful only if they are successfully decoded within τ time slots after their arrival. Next, we differentiate flows according to their *relative priority* and their *long-term reliability requirements*. To model their relative priority, we assign a set of positive weights $\mathbf{w} = (w_f)f$ to each flow, and try to maximize the weighted sum of their mean flow rates. To model their long-term reliability requirements, we use $\mathbf{q} = (q_f)_f$, where $q_f \in [0, 1)$ imposes an upper bound on the *fraction of dropped Flow-f packets in the long run*. In other words, we are forced to guarantee that

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{D_f[t]}{A_f[t]} \quad \leq \quad q_f, \quad \forall f, \tag{2}$$

with the convention that $0/0 = 0$. Hence, it is required that at least $(1 - q_f)$ fraction of arriving Flow-f packets be served within their deadline on average. Thus, the requirement gets more stringent as q_f decreases towards 0. Real-time applications such as voice/video transfers that can tolerate a certain fraction of packet losses typically have such *delivery ratio requirements*. This traffic modeling follows that of [7], [8], and is attractive for both practical modeling and theoretical analysis purposes.

Qualitative Description of the Objective: For the generic communication system of Figure 1, we aim to design a joint rate controller and dynamic coding strategy that maximizes $\mathbf{w}-$weighted sum of generated flow rates subject to the short-term deadline constraint τ of all packets and the long-term reliability requirements $\mathbf{q}$ of the flows given in (2). In other words, given the weights $\mathbf{w}$ and reliability requirements $\mathbf{q}$, our objective is to push the flow rates to the limit of what is supportable by the given communication medium.

In the following sections, we will formulate this problem mathematically and provide a stochastic algorithm to solve it. This algorithm is also applied to an important downlink broadcasting setup.

III. Stochastic Control Problem Formulation

In this section, we provide an *infinite-horizon optimal stochastic control* formulation of the qualitative objective presented in Section II, namely that of joint rate allocation and coding for maximizing a given weighted sum of flow rates while satisfying both the short-term deadline requirements and the long-term reliability requirements (as described in (2)) of the flows.

For ease of exposition, we define expressions for the long-term flow rates and successful service rates as:

$$\overline{\lambda}_f := \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[A_f[t]], \;\; \underline{\lambda}_f := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[A_f[t]],$$

and $\underline{\mu}_f := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[M_f[t]]$, where the expectations are over the randomness of the packet generation process and the completion times of coding blocks, which are, in turn, functions of the coding strategy and the wireless channel variations. We are now ready to pose our stochastic control problem (SCP) as:

Definition 1 ($\infty-$Horizon Stochastic Control Problem):

$$\begin{aligned} \text{(SCP):} \quad & \underset{\{\mathbf{A}[t],\mathbb{K}[t]\}_{t\geq 1}}{\text{Maximize}} \quad \sum_f w_f \underline{\lambda}_f \\ & \text{subject to} \quad K^f[t] \leq A_f[t], \quad \forall f, \forall t \geq 1, \quad (3) \\ & \qquad \overline{\lambda}_f(1 - q_f) \leq \underline{\mu}_f, \quad \forall f, \quad (4) \\ & \qquad \mathbf{M}[t] \sim F_{g(\mathbf{Z}(\mathbb{K}[t]))}(m), \;\; \forall t \geq 1, (5) \end{aligned}$$

where (3) assures that the coding is limited to packets available in that frame, (4) assures that long-term delivery ratio requirements are satisfied, and (5) indicates that the successful packet transmissions are random as a function of the completion-time distributions and coding decisions.

We emphasize that solving SCP through standard control theoretic methods is extremely difficult, and likely impossible. Not only is it an infinite-horizon problem, but also it contains instantaneous constraints (3) and randomness described by convoluted distributions (5) as well as long-term average requirements (4). In our work, instead of pursuing the solution

of this complex problem through standard stochastic controller design, we utilize a combination of *dynamic pricing* and *finite-horizon dynamic programming* strategies that yields a practically implementable solution.

IV. Algorithm Description and Analysis

In this section, we propose a practical stochastic algorithm that solves SCP (see Definition 1). Essentially, the algorithm introduces and maintains a time-varying price vector $\mathbf{X}[t] = (X_f[t])_f$, where $X_f[t]$ measures the experienced reliability requirement violation for Flow f. Then, the rate controller and the dynamic coding strategy uses $\mathbf{X}[t]$ to determine the mean number of newly generated packets and the composition $\mathbb{K}[t]$ of the coding window selection in frame t.

Definition 2 (Joint Rate Control and Coding Algorithm): For a given set $\mathcal{F}$ of τ-deadline-constrained flows, their relative weights $\mathbf{w} = (w_f)$, and their long-term reliability requirements $\mathbf{q} = (q_f)$, the joint dynamic algorithm performs the following operations:

► **Price Update**: We maintain a price variable $\mathbf{X}[t] = (X_f[t])_f$, where $\{X_f[t]\}_t$ for each f is initiated at $x_f[1] = 0$ and is updated at each frame according to:

$$X_f[t] = (X_f[t-1] + \beta(D_f[t-1] - q_f A_f[t-1]))^+ , \quad (6)$$

where $(y)^+ = max(0, y)$, and $\beta > 0$ is a small stepsize parameter. We recall that $A_f[t-1]$ and $D_f[t-1]$ denotes the number of generated and dropped Flow-f packets in frame (t-1), and hence are known at the beginning of frame t.

► **Rate Controller**: Given $\mathbf{X}[t]$, the rate controller updates the vector $\lambda[t] = (\lambda_f[t])_f$ of mean number of packets to be generated in frame t as:

$$\lambda_f[t] = (\lambda_f[t-1] + \alpha(w_f - X_f[t](1-q_f)))^+ , \quad \forall f, \quad (7)$$

where $\alpha > 0$ is a small parameter. Then, a random vector, $\mathbf{A}[t]$, of packets are generated with the computed mean vector $\lambda[t]$ and with a finite variance[2].

► **Dynamic Coding Strategy**: Given $\mathbf{X}[t]$ and $\mathbf{A}[t]$, the dynamic encoder selects the coding block sizes and composition for frame t, namely $\mathbb{K}[t]$, such that

$$\mathbb{K}[t] \quad \in \quad \underset{\{\mathbb{K}\in\mathcal{K}: K^f[t]=A_f[t], \forall f\}}{\arg\max} \sum_f X_f[t]\mathbb{E}[M_f(\mathbb{K})], \quad (8)$$

where $\mathcal{K}$ denotes the set of all possible coding matrix choices, and $\mathbb{E}[M_f(\mathbb{K})]$ compactly represents the expected number of successfully decoded Flow-f packets in the frame when the coding decision is $\mathbb{K}$.

The decision problem in (8), in turn, is solved through *Finite-Horizon Dynamic Programming* by defining the optimal reward-to-go function $J^\star(\mathbf{B}, s)$ as the maximum value of the $\mathbf{X}[t]-$weighted mean success rates in (8) when there are a vector of $\mathbf{B} = (B_f)$ packets waiting for transmission while there are $s \in \{0, \cdots, \tau\}$ slots left until the end of the frame. Then, $J^\star(\mathbf{B}, s)$ satisfies Bellman's equation ([1]):

$$J^\star(\mathbf{B}, s) = \max_{\{\mathbf{K}_1: K_{f,1} \le B_f, \forall f\}} \{\mathbb{E}[J^\star((\mathbf{B} - \mathbf{K}_1), (s - Z(K_1))) + (\sum_f K_{f,1} X_f[t]) \cdot \mathcal{I}(Z(K_1) \le s)]\},$$

where $\mathcal{I}(B)$ is the indicator function of event B.. This is solved through backward recursion with the initial conditions: $J^\star(\mathbf{B}, 0) = 0$, for all $\mathbf{B}$. The solution also yields the optimal choice $\mathbf{K}^\star(\mathbf{B}, s)$ of the first coding window size for all $\mathbf{B}$ and s of interest. Then, the dynamic coding strategy is to select $\mathbf{K}^\star(\mathbf{A}[t], \tau)$ as the first coding window composition. If the transmission of that block completes within the frame, leaving s more slots until the end of the frame, the dynamic controller re-uses to the same DP solution (without any recomputing) to determine the optimal size of the second coding window for the remaining packets and time, and so on.

We note that the price update rule (6) uses an *instantaneous* measure of the long-term fractional packet drop requirement to change $X_f[t]$ with increments of β. The key observation is that if $\{X_f[t]\}_t$ is guaranteed to be stable in the long run, the long-term delivery ration requirement will be met. In fact, the rate controller and the coding policy both aim at controlling the arrival and service rates to guarantee the stability of $\{\mathbf{X}[t]\}_t$. Accordingly, the rate controller of (7) discourages packet generation for flows if their scaled price value $X_f[t]$ exceeds their weights. Also, the coding strategy of (8) weighs the successful service rates of flows with their current prices, $\{X_f[t]\}_f$, therefore effectively prioritizing the service of those flows whose reliability requirement is violated more severely. Finally, notice that the rate controller includes a random mapping between the mean rates $\lambda[t]$ and the actually generated packets $\mathbf{A}[t]$ to render the rate controller more practical by allowing for stochastic variations around the desired mean values.

We provide a description and an analysis of a deterministic fluid approximation of this stochastic algorithm in I, which establishes that the fluid approximation is globally asymptotically stable ([9], [12]), and that it converges to the optimum solution of the deterministic formulation of (SCP). It is well-known (e.g. see [4], [6], [11]) that such fluid analysis provides the foundation for proving the optimality of its stochastic counterpart, which is part of our future work.

V. Application: Cellular Downlink Scenario

The generic model of the communication system can be used for lots of specific communication scenarios and coding strategies. In this section, we describe an important example whereby a Base Station (BS) is serving multiple flows by broadcasting their incoming packets to N receivers over time varying channels (see Figure 3). Each packet is a vector of length l over a finite field $\mathbb{F}_d$, for some $d \in \mathcal{Z}_+$.

All the components of this particular system follows the descriptions provided in Section II-A except that we can now express the completion time $Z(K)$ more precisely for different coding strategies, in particular *random network coding* and *scheduling* strategies that are used for comparison purposes.

[2]For example, $A_f[t]$ may be uniform, two-point Bernoulli, or Poisson distributed with the updated mean $\lambda_f[t]$.

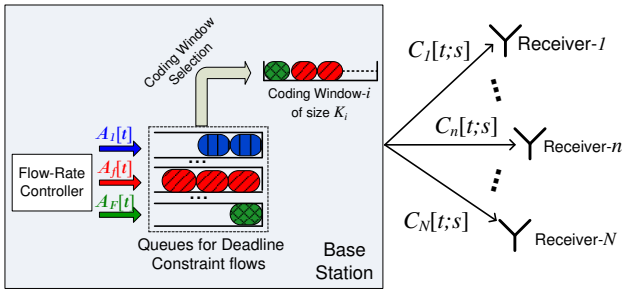

Fig. 3. Cellular downlink model for broadcasting deadline-constrained flows to N users over erasure channels.

To this end, we assume that the channel conditions of each receiver independently vary in each time slot. Specifically, let $C_n[t;s]$ denote the state of the nth receiver's channel in the sth time slot, with $s \in \{1, \cdots, \tau\}$, of the tth frame. We assume that $C_n[t;s]$ is an independent Bernoulli process with $c_n = P(C_n[t;s] = 1) = 1 - P(C_n[t;s])$. If $C_n[t;s] = 1$ in a time slot, then the packet transmitted by the BS is assumed to be successfully received by Receiver-n. The BS is assumed to know (c_n) but not the realizations before its transmissions. Note that in this setup, the completion time of a block of K packets is a function of K and the number of receivers N being served. To highlight the dependence on N, we use $Z(K,N)$ to denote the completion time of the block.

Next, we describe two key strategies that the BS can employ to transmit its block of packets.

Definition 3 (Randomized Broadcast Coding (RBC)): A network coding strategy over a block of K packets where in a slot, say s; any linear combination of the K packets in the file can be transmitted. Specifically if $\mathbf{P}(s)$ denote the packet chosen for transmission in slot s, we have $\mathbf{P}[s] = \sum_{k=1}^{K} \alpha_k[s] P_k$, where $\{\alpha_k[s]\}_k$ are chosen uniformly at random from the field $\mathbb{F}_d \backslash \{0\}$ for every time slot $[s]$.
It has been shown in [5] that RBC is an optimal coding strategy as the field size $d \to \infty$. Also, the distribution of the completion time $F_{Z^{RBC}(N,K)}$ [3] is given as

$$F_{Z^{RBC}(N,K)}(x) = (1 - h(x)(1-c)^x)^N \tag{9}$$

where $h(x) = \beta x^{(K-1)}$ with $\beta := (\frac{c}{1-c})^{K-1} \frac{1}{(K-1)!}$.

Definition 4 (Round Robin Scheduling (RR)): For a given block of packets of size K, the BS at any given slot broadcasts a single packet from the current coding window. Thus, we have $\mathbf{P}(s) \in \{\mathbf{P}_k\}_{\{k=1,\ldots,K\}}$. In the optimal Round Robin Scheduling (see [5] for proof of optimality) Packet-k is transmitted in time-slots $(rK + k)$ for $r = 0, 1, \ldots$ until all the receivers receive the whole block.

Proposition 1: The distribution of the completion time $F_{Z^{RR}(N,K)}$ for the RR strategy [3] is given by

$$F_{Z^{RR}(N,K)}(x) = \sum_{y=0}^{x} \Bigg(\underbrace{(1-(1-c)^{r+1})^{N(k-1)}}_{(a)} \cdot \underbrace{(1-(1-c)^r)^{N(K-k)}}_{(b)} \cdot \underbrace{\Bigg(\sum_{n=1}^{N} \binom{N}{n} (1-(1-c)^r)^{N-n} ((1-c)^r c)^n \Bigg)}_{(c)} \Bigg).$$

[3]Here the expressions for $F_{Z^{RBC}(N,K)}$ and $F_{Z^{RR}(N,K)}$ are provided for symmetric channel conditions for ease of exposition, but can also be expressed for asymmetric conditions.

where $r = \lfloor (y/K) \rfloor$ and $k = y \bmod (K)$ when $y \bmod (K) \neq 0$; $r = (y/K) - 1$ and $k = K$ when $y \bmod (K) = 0$.

Proof: The proof can be found in the Appendix II. ■

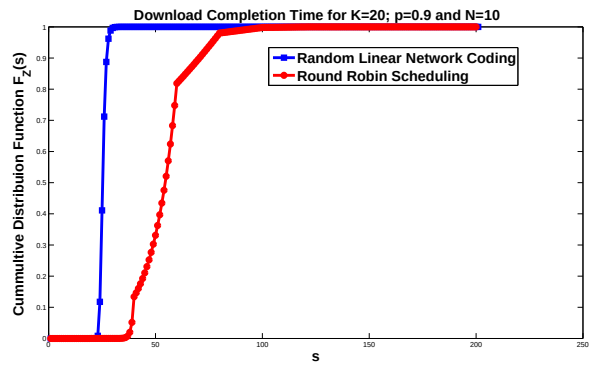

Fig. 4. Download Completion Time Distribution

It is known that ([5]) that the expected completion time for RBC strategy is lower than that for RR strategy and the difference grows as the number of receivers N and the block size K increase when there are no deadline constraints. Figure 4 illustrates the download completion time distributions of RBC and RR strategies. In this work we extend the comparison to the case when the packets have strict deadline constraints and a quality metric to satisfy in the form of reliability requirement.

VI. Simulations

In this section, we provide simulation results to complement the analysis in the previous sections. The purpose of the simulation is to understand the effect of various parameters on the algorithm and how the joint rate controller and dynamic coding strategy behaves under different conditions. The simulations are presented for the cellular downlink application introduced in the section V. We make the reasonable assumption that the channels between the base station and each of the users are independent and identical in distribution. We let the arrival distribution to be uniform with domain 0 to $2\lambda_f$ and mean λ_f and set the frame size of $\tau = 3$ in most of the simulations. We consider two equally prioritized ($\mathbf{w} = [2;2]$) flows in most our simulations which also are identical in their long term reliability requirements.

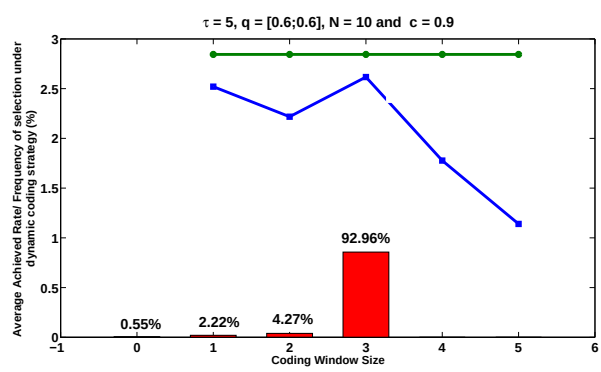

Fig. 5. Fixed Coding Window vs Dynamic Coding Window

Fixed vs. Dynamic Coding Window: Figure 5 compares the developed dynamic coding algorithm to a strategy where the coding window size is kept constant.The figure indicates that the fixed coding window of 3 performs the best among the various fixed coding window sizes. The dynamic algorithm also chooses $K = 3$ most frequently (92.96%) but also dynamically adapts its window size, when the time to deadline is decreasing, thereby outperforming the fixed window strategy.
Effect of the Transmission Strategy and N: In figure 6 we plot the achieved rates for two flows with variations in the number of receivers in the cellular downlink system and the transmission strategies. It can be seen that the achieved rates are lower when the number of receivers is increased and also when the strategy is changed from RBC to RR. Both these can be attributed to the larger expected completion times.

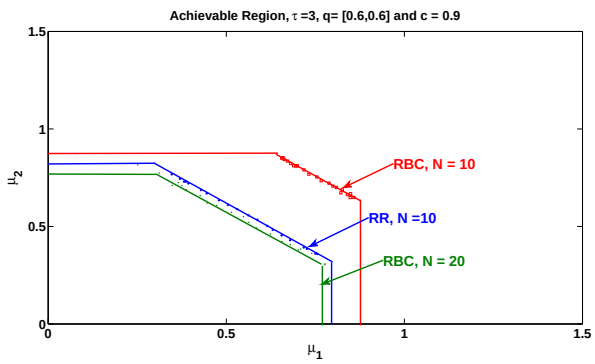

Fig. 6. Achieved Rates for varied number of receivers and transmission strategies

Effect of Reliability Metric, q: The figure 7 illustrates the effects of long term reliability metric **q** with both RBC and RR strategies. It indicates the failure of the strategies to serve the flows with very high demands ($q_f < 0.55$), while providing the maximum achievable rate to the lower demand flows ($q_f > 0.9$). The lower expected completion time of the RBC strategy allows it to serve flows with more stringent reliability requirement ($0.55 \geq q_f$) at a higher rate than the RR strategy($0.75 \geq q_f$).

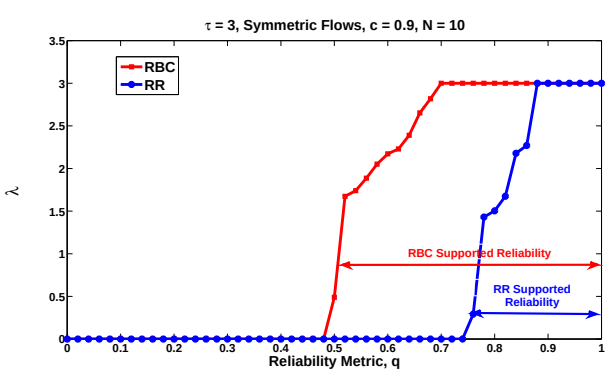

Fig. 7. Achieved Rates for RBC and RR Strategies

The figure 8 gives the comparison for the flows with unsymmetrical reliabilities. The reliability of one of the flows is fixed while that of other is changed over a range. It can be seen that when the more stringent flow (Flow-1) is unsupported ($q_f \leq 0.5$), then the system provides the maximum achievable rate to Flow-2. While when both the flows are relatively lax($q_f > 0.8$) the achievable rates are increased to the maximum, but the system prefers the more stringent flow and hence the average service is higher for Flow-2.

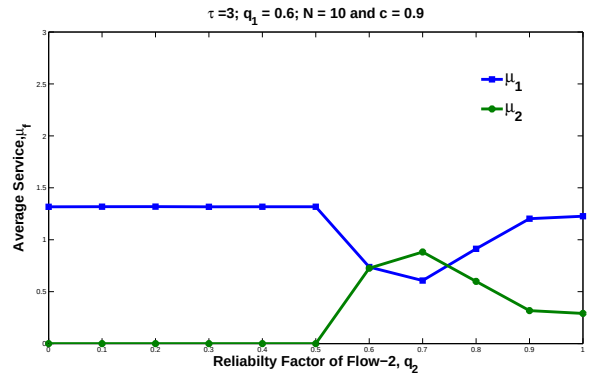

Fig. 8. Asymmetric Reliability Requirement Scenario

VII. Conclusion

In this paper we have presented an optimal rate control and coding operation for deadline constrained traffic over time varying wireless channels. A model was developed for a general communication system with heterogeneous reliability requirements and priorities of different deadline constrained flows. A novel algorithm operating at different timescales was presented that combines dynamic pricing to determine the incoming flow rates and finite horizon dynamic programming for coding operations. Analysis of the continuous-time fluid approximation of the stochastic system operating under the dynamic algorithm was also provided. The developed generic algorithm is also applied to the important scenario of cellular broadcast over erasure channels with random network coding. Simulations corroborate our results and also show the effect of various system parameters and application requirements.

Appendix I
Fluid Model Analysis

This section is composed of three items: the formulation of a static optimization problem that is equivalent to SCP of Definition 1; the fluid modeling of the stochastic algorithm provided in Definition 2; and finally the analysis of the resulting deterministic system operation under the fluid approximation of the stochastic algorithm.
1. Deterministic Optimization Formulation: Let us define Ω as the set of mean successful deadline-constrained service rates that the wireless channel(s) can support under all possible coding window selection strategies. In other words, this static region, denoted by Ω, contains vectors $\mu = (\mu_f)_f$ of mean rates, where μ_f gives the mean service rate (in packets per frame) that Flow-f packets receives within τ slots of their arrival, under some feasible coding policy, say π.

We note that our subsequent discussion does not require the exact description of Ω, which is cumbersome. Yet, we remark that this region must be convex since any rate that is a convex combination of two achievable rates can be achieved by appropriately time sharing between the corresponding two policies. The abstract definition of this region allows us to

formulate a Deterministic Optimization Problem (DOP) that is loosely equivalent to SCP.

Definition 5 (Deterministic Optimization Problem):

$$\text{(DOP):} \quad \underset{\{\lambda\geq\mathbf{0},\mu\geq\mathbf{0}\}}{\text{Maximize}} \quad \sum_f w_f\lambda_f$$

$$\text{subject to} \quad \lambda_f(1-q_f)\leq\mu_f, \quad \forall f, \tag{10}$$

$$\mu\in\Omega. \tag{11}$$

Notice that (10) is equivalent to (4), and (11) captures the restrictions (5) through its Ω region.

2. Fluid Approximation of the Stochastic Joint Algorithm: The operation of the stochastic algorithm of Definition 2 is approximated by the differential equations and decisions as follows.

Definition 6 (Fluid Model of Joint Algorithm): We assume that $\lambda(t)$, $\mu(t)$, and the price variable $\chi(t)$ evolve in continuous-time (as indicated by the use of (t) instead of [t] in the notation) according to the following rules:

Price Update: The price variable $\chi(t)=(\chi_f(t))_f$ evolves according to the differential equation:

$$\dot{\chi}_f(t):=\frac{d\chi(t)}{dt}=\beta\left((1-q_f)\lambda_f(t)-\mu_f(t)\right)\right)^+_{\chi_f(t)}, \tag{12}$$

where $(y)^+_z=y\mathcal{I}(z>0)+max(y,0)\mathcal{I}(z=0)$, and $\beta>0$ as before. Note that (12) is the differential equation version of the difference equation (6) when we note that $\mathbf{D}[t]=\mathbf{A}[t]-\mathbf{M}[t]$.

Rate Controller: Given $\chi(t)$, the rate controller updates the vector $\lambda(t)=(\lambda_f(t))_f$ of mean number of packets to be generated at time t according to:

$$\dot{\lambda}_f(t)=\alpha\left(w_f-\chi_f(t)(1-q_f)\right)^+_{\lambda_f(t)}, \quad \forall f, \tag{13}$$

where $\alpha>0$ is the same small parameter in (7).

Dynamic Coding Strategy: Given $\chi(t)$, the dynamic encoder selects the policy that yields the deadline-constrained mean service rate vector $\mu(t)$ that satisfies

$$\mu(t) \quad\in\quad \underset{\eta\in\Omega}{\arg\max}\sum_f\chi_f(t)\eta_f. \tag{14}$$

We note the similarity of the objectives of this service decision rule and (8). We also remark that the solution of (14) is practically impossible, while we already showed that (8) is solvable through finite-horizon dynamic programming. Yet, (14) will be useful for analysis purposes.

3. Analysis of the overall deterministic fluid model: It is easy to see that the DOP in Definition 5) is a convex problem that satisfies the Slater's condition, and hence has no duality gap. We let χ_f denote the *dual variable* (or *Lagrange multiplier*) associated with the fth reliability constraint in (10). Then, there exists a set, $\{(\lambda^\star,\mu^\star,\chi^\star)\}$ of optimal primal-dual variables satisfying the KKT conditions ([3]):

(i) $0\leq\lambda^\star_f(1-q_f)\leq\mu^\star_f$, for all f.

(ii) $\mu^\star\in\underset{\eta\in\Omega}{\arg\max}\sum_f\chi^\star_f\eta_f$.

(iii) $\lambda^\star\in\underset{\gamma\geq0}{\arg\max}\sum_f\gamma_f(w_f-\chi^\star_f(1-q_f))$.

(iv) $\chi^\star_f(\lambda^\star_f(1-q_f)-\mu^\star_f)=0$, for all f.

The next theorem proves that the fluid model of Definition 6 converges to the optimal solution set.

Theorem 1 (Global Optimality of the Fluid Approximation): Starting from any feasible initial state $(\lambda(0),\mu(0),\chi(0))$, the fluid algorithm described in Definition 6 converges to the set $\{(\lambda^\star,\mu^\star,\chi^\star)\}$ of optimal primal-dual solutions of the DOP described in Definition 5. Thus, we call is *global asymptotically optimal*.

Proof: For any primal-dual optimal solution $(\lambda^\star,\mu^\star,\chi^\star)$, consider the Lyapunov function:

$$W(\lambda,\mu,\chi) \quad=\quad \frac{1}{2}\sum_f\left(\frac{(\lambda_f-\lambda^\star_f)^2}{\alpha}+\frac{(\chi_f-\chi^\star_f)^2}{\beta}\right).$$

We study the time-derivative of this function under the operation of the fluid algorithm that evolves as in (12)-(14).

$$\begin{aligned}
&\dot{W}(\lambda(t),\mu(t),\chi(t))\\
&= \sum_f(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))^+_{\lambda_f(t)}\\
&\quad+\sum_f(\chi_f(t)-\chi^\star_f)((1-q_f)\lambda_f(t)-\mu_f(t))^+_{\chi_f(t)}\\
&\overset{(a)}{\leq} \sum_f(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))\\
&\quad+\sum_f(\chi_f(t)-\chi^\star_f)((1-q_f)\lambda_f(t)-\mu_f(t))\\
&\overset{(b)}{=} \sum_f(\lambda_f(t)-\lambda^\star_f)(w_f-\chi^\star_f(1-q_f)) && (15)\\
&\quad+\sum_f(\lambda_f(t)-\lambda^\star_f)(1-q_f)(\chi^\star_f-\chi_f(t)) && (16)\\
&\quad+\sum_f(\chi_f(t)-\chi^\star_f)((1-q_f)\lambda^\star_f-\mu^\star_f) && (17)\\
&\quad+\sum_f(\chi_f(t)-\chi^\star_f)(1-q_f)(\lambda_f(t)-\lambda^\star_f) && (18)\\
&\quad+\sum_f(\chi_f(t)-\chi^\star_f)(\mu^\star-\mu_f(t)). && (19)
\end{aligned}$$

To see the rationale behind the inequality (a) consider the term $(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))^+_{\lambda_f(t)}$. Note that when $\lambda_f(t)>0$, we have $(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))^+_{\lambda_f(t)}=(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))$ by definition. On the other hand, when $\lambda_f(t)=0$, $(w_f-\chi_f(t)(1-q_f))^+_{\lambda_f(t)}\geq0$, and hence $(\lambda_f(t)-\lambda^\star_f)(w_f-\chi_f(t)(1-q_f))^+_{\lambda_f(t)}\leq0$. This shows the inequality (a) for this term, and a similar argument holds for all other summands.

To get the equality (b), we have added and subtracted the following terms and re-arranged them:

$$\sum_f[(\lambda_f(t)-\lambda^\star_f)\chi^\star_f(1-q_f)+(\chi_f(t)-\chi^\star_f)(\lambda^\star_f(1-q_f)-\mu^\star_f)].$$

Notice that (16) and (18) of the sum cancel each other. Also, note that the KKT condition (iii) implies that $\lambda^\star_f(w_f-\chi^\star_f(1-q_f))=0$, for all f. This, in turn, implies that (15) $\leq$ 0 with strict inequality if $\lambda(t)\neq\lambda^\star$. Similarly, the KKT condition (iv) can be used in (17) to show that (17) $\leq$ 0 with strict inequality if $\chi(t)\neq\chi^\star$. Finally, we consider (19): recall that $\mu(t)\in\arg\max_{\eta\in\Omega}\sum_f\chi_f(t)\eta_f$ and

$\mu^\star \in \arg\max_{\eta\in\Omega} \sum_f \chi_f^\star \eta_f$. These imply that $\sum_f [\chi_f(t)(\mu_f^\star - \mu_f(t))] \leq 0$ and $\sum_f [\chi_f^\star(\mu_f(t) - \mu_f^\star)] \leq 0$. Thus, we must have $(19) \leq 0$, with strict inequality when $\mu(t) \neq \mu^\star$. Combining all these results yield $\dot{W}(\lambda(t), \mu(t), \chi(t)) \leq 0$ with strict inequality when $(\lambda(t), \mu(t), \chi(t)) \neq (\lambda^\star, \mu^\star, \chi^\star)$, which, by Lyapunov Stability Theorem ([9]) yields global asymptotic optimality of the fluid algorithm. ■

APPENDIX II
PROOF OF PROPOSITION 1

Proof: If X_k^n is defined to be the number of transmissions of $\mathbf{P}_k$ before it is received by Receiver-i. Then, $Y^n = \max_{k\in\{1,\ldots,K\}}\{KX_k^n + k\}$ gives the time slot when Receiver-n receives the whole file. Finally, $Z^{RR}(N,K) = \max_{n\in\{1,\ldots,N\}} Y^n$ gives the download completion time of the Round-Robin scheduler. The terms expressed in the proposition as (a), (b) and (c) are

(a) *Pr*(Packet $j \in \{1,\ldots,k-1\}$ received successfully by all N users) = *Pr*(At least one ON channel in $s+1$ transmission opportunities.) = $(1-(1-c)^{s+1})^{N(k-1)}$.

(b) *Pr*(Packet $j \in \{k+1,\ldots,K\}$ received successfully by all N users) = *Pr*(At least one ON channel in s transmission opportunities.) = $(1-(1-c)^{s})^{N(k-1)}$.

For packet k n of the users can complete in slot m while the others might have completed in any of the previous s transmissions of packet k. Hence,

(c) *Pr*(Packet k received successfully by all N users in slot m) = $\sum_{n=1}^{N}\binom{N}{n}(1-(1-c)^s)^{N-n}((1-c)^s c)^n)$.

Putting the three (a),(b) and (c) together we obtain the result. ■

REFERENCES

[1] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[2] P.P. Bhattacharya and A. Ephremides. Optimal scheduling with strict deadlines. *Automatic Control, IEEE Transactions on*, 34(7):721–728, Jul 1989.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[4] J. G. Dai and S. P. Meyn. Convergence and moments of networks and their fluid models. *IEEE Transactions on Automatic Control*, 40:1889–1904, 1995.

[5] A. Eryilmaz, A. Ozdaglar, and M. Medard. On delay performance gains from network coding. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 864–870, March 2006.

[6] 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.

[7] I. H. Hou, V. Borkar, and P. R. Kumar. A theory of qos for wireless. *INFOCOM 2009, IEEE*, pages 486–494, June 2009.

[8] J. J. Jaramillo and R. Srikant. Optimal Scheduling for Fair Resource Allocation in Ad Hoc Networks with Elastic and Inelastic Traffic. *ArXiv e-prints*, July 2009.

[9] H. Khalil. *Nonlinear Systems*. 3rd ed., Prentice Hall, Upper Saddle River, NJ, 2002.

[10] D.G. Pandelis and D. Teneketzis. Stochastic scheduling in priority queues with strict deadlines. *Advances in Applied Probability*, 26:258–279, 1994.

[11] Srinivas Shakkottai and R. Srikant. Network optimization and control. *Found. Trends Netw.*, 2(3):271–379, 2007.

[12] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhäuser, Boston, MA, 2004.

[13] Z.-X. Zhao, S.S. Panwar, and D. Towsley. Queueing performance with impatient customers. In *INFOCOM '91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s., IEEE*, pages 400–409 vol.1, Apr 1991.