Economic Aspects of Network Coding

Ebad Ahmed, Atilla Eryilmaz, Asuman Ozdaglar and Muriel Médard

***Abstract*— This work focuses on a cellular system scenario of a single base station broadcasting incoming files (or users) to multiple receivers over a time-varying channel. The base-station sets a price per receiver and per file to maximize its profit. The files may or may not enter the system depending on the price of service, the delay performance and their randomly determined valuations.**

Under this scenario, we consider the use of *Network Coding* and *Scheduling* as the transmission strategies. Our analysis provides approximate characterization of the optimal admission rate, price and revenue as functions of the first and second moments of the service time processes under mild assumptions. We show that optimal network coding window size is highly insensitive to the number of receivers, which suggests that pricing and coding decisions can be decoupled. Moreover, we demonstrate that network coding leads to a significant (almost ten-fold) gain in the revenue compared to scheduling.

I. Introduction

The advent of wireless networks has revolutionized the whole domain of packet networks. The arrival of cellular systems has resulted in an increased demand for wireless networks for transmitting data, and has also expedited the growth of the capacity of wireless networks. Wireless networks are employed for a variety of purposes, including transmission of data in multicast settings such as video-conferencing and file transfer.

The traditional routing approaches, which work reasonably well in wireline networks, fail to utilize the full capacity of such wireless networks because of unreliable links and packet losses. In light of the inefficacy of the hop-by-hop routing approach, it has been recognized that broadcasting to multiple destinations may be accomplished much more efficiently if *network coding* is used (cf. [2], [9], [10]). Network coding has been employed in a multitude of areas, such as network management, overlay networks, and wireless networks (e.g. [7], [15], [8], [5]). In addition to providing an immediate increase in the capacity of the network over traditional routing schemes, network coding also allows efficient computation of a minimum-cost subgraph for a single multicast session given a fixed (i.e., inelastic) rate demand, where the cost is defined as the sum of the link costs [11]. More recent work [6] developed a model to study delays in file downloads with network coding and quantified the gains resulting from network coding relative to the traditional scheduling methods.

This work was supported by DARPA CBMANET contract BAA-05-42.

All authors are with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA,02139, USA {ebad, eryilmaz, asuman, medard}@mit.edu

The objective of this paper is to analyze the economic gains obtained from network coding in a dynamic *rateless* transmission scenario[1]. We consider the cellular downlink transmission of files from a single base station to multiple receivers. Our model involves a stream of incoming files (users), which arrive according to a Poisson process and are broadcast to all the receivers. The base station sets the entry price per receiver and per user with the goal of maximizing its revenue, recognizing the effect that his decision will have on the choice of users to enter the network. We assume that each user derives a utility from the file transfer to multiple receivers, while incurring delay and monetary costs. We are interested in obtaining explicit characterizations of the maximum revenue attained by the base station, the optimal user admission rate, and the optimal price charged by the base station. In addition, we analyze gains in the revenue when the base station chooses to use network coding for file transmission instead of traditional scheduling schemes.

The rateless transmission model that we use has been introduced in [6], where the mean file completion times of network coding and scheduling strategies have been studied. In that work, significant gains have been observed, both with and without channel side information, in the performance of network coding as opposed to scheduling. Our work differs from [6] in that we consider the benefits obtained from network coding from a profit-maximizing service provider viewpoint. Our analysis shows how the delay gains translate into economic benefits in a dynamic setting. In this analysis, not only the first moment but also the second moment of the service time process influence the performance.

There have been a few other works that studied the economic aspects of queueing systems (e.g. [13], [12], [1]). However, none of these works consider the broadcasting scenario with the possibility of network coding. The non-standard service time distributions of the network coding and scheduling strategies prevent the use of earlier analysis and results for our problem. Moreover, we are interested in the effect of practical parameters such as the coding window size and the number of receivers on the system performance.

Our analysis reveals a number of interesting results: We show under mild assumptions that the revenue is a unimodal function of the file size. Furthermore, if the base station operates under low-traffic conditions and the number of receivers is assumed to be large, the file size that maximizes the revenue is almost completely insensitive to changes in the number of receivers. This surprising property allows for

[1] In a rateless transmission scenario, the focus is on the transmission of chunks containing a fixed amount of data such as files, rather than a constant flow of data.

the decoupling of the optimal coding window size from the pricing decision of the base station. We show through simulations that network coding provides significant gains in revenue compared to scheduling, while maintaining a high user admission rate. We also study the effect of channel conditions on the revenue of the base station.

The rest of the paper is organized as follows: In Section II, we describe the model for the transmission of files to multiple receivers. In Section III, we set up the revenue maximization problem and obtain a characterization of the optimal admission rate through the use of extreme value theory. In Section IV, we use our characterization of the optimal admission rate to obtain approximate expressions for the optimal price and the optimal revenue in terms of the first and second moments of the service time at the base station, and show that the optimum file size is insensitive to variations in the number of receivers. In Section VI, we analyze through simulations the gains in revenue when the base station uses network coding as opposed to scheduling. In Section VII we conclude with final remarks and ideas for future work.

II. System Model

We consider the downlink of a base station broadcasting a sequence of incoming files (users) to N receivers over time varying channels. Upon arrival to the system, each user decides whether it will enter the queue for service based on its valuation of the service and the price charged by the service provider. The files that have entered the queue are served in a First-In-First-Out (FIFO) fashion. Thus, the transmission of the next file starts after the current file has been received by every receiver.

Each file is assumed to be composed of K *packets*, where Packet-k of a given file is referred to as $\mathbf{P}_k$, which is a vector of length m over a finite field $\mathbb{F}_q$, for some $q \in \mathcal{Z}_+$. Transmissions take place in regularly arranged time slots with each slot long enough to accommodate a single packet transmission. The channel between the base station and each receiver has a time varying nature to capture the influence of changing channel conditions, possible interference effects and the mobility of the receivers. Specifically, we assume that the channel condition in slot t between the base station and the n^{th} receiver is captured by a Bernoulli distributed random variable $C_n[t]$ with mean c_n that is independent across users and time slots. When $C_n[t] = 1$, the channel is assumed to be ON and the transmission of the base station is successfully received by the n^{th} receiver. If, on the other hand, $C_n[t] = 0$, the transmitted packet does not reach receiver n. We will refer to c_n as the mean channel rate for channel n.

Let us use $\mathbf{P}[t]$ to denote the packet chosen for transmission in slot t. If the base station is not allowed to code, then at any given slot it must transmit a single packet from one of the files. Thus, we have $\mathbf{P}[t] \in \{\mathbf{P}_k\}_{\{k=1,\cdots,K\}}$. This is the typical mode of transmission considered in literature. We will refer to this mode as the *Scheduling Mode* (or simply *Scheduling*).

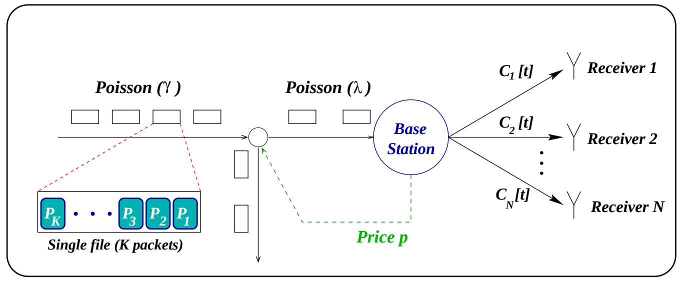

Fig. 1. System model.

If coding is allowed, then in a slot, say t, any linear combination of the packets can be transmitted. Specifically, we have

$$\mathbf{P}[t] = \sum_{k=1}^{K} a_k[t] \mathbf{P}_k,$$

where $a_k[t] \in \mathbb{F}_q$ for each $k \in \{1, \cdots, K\}$. The transmitter chooses the coefficients $\{a_k[t]\}$ at every time slot t. This mode of transmission will henceforth be referred to as the *Coding Mode* (or simply *Coding*).

The strategy employed by the base station to broadcast the head-of-the-line file to the receivers has a critical effect on the service time distribution of the base station. In [6], an extensive analysis of the delay performance of such a file download is provided under Network Coding/Scheduling, and the presence/lack of Channel-Side-Information (CSI). The authors propose a randomized coding strategy, where at every time slot $a_k[t]$ is chosen uniformly at random from $\mathbb{F}_q \backslash \{0\}$. It is shown that such a policy is delay optimal both in the presence and lack of CSI. Unless N is very small, the assumption of the availability of CSI is impractical because of the requirements of frequent feedback and training signals. Therefore we focus on the realistic scenario where no CSI is available at the base station, and feedback is sent only when a receiver gets the whole file. Such a system is not only simpler to implement, but also dissipates less energy and bandwidth resources.

Throughout, we will use the terms *file* and *user* interchangeably. We assume that users arrive according to a Poisson process of rate $\gamma > 0$ to be broadcast to all the receivers. The base station charges each user a price p per receiver for the file transfer. Thus, broadcasting to N receivers costs a total amount Np to the user. Each user has the option of either accepting or refusing the services provided by the base station. This model is depicted in Figure 1. The decision is based on the utility derived by the user on accepting the service, the delay it will experience before the completion of the download, and the price it will pay to the base station. In particular, we assume that each user will derive a utility $U^{(N,K)}$ from transferring a single file of size K to the N receivers, where $U^{(N,K)}$ is a random variable with distribution function $F_{U^{(N,K)}}(\cdot)$.

The system can be effectively modeled as an $M/G/1$ queue, and each user will experience a delay $D(\gamma, p, N)$ depending on the transmission strategy used by the base-station (network coding or scheduling) and the number of users waiting in the queue (dictated by the arrival rate γ). The expression for the expected delay is given by the celebrated *Pollaczek-Khinchin* formula:

$$\mathbb{E}[Delay] = \frac{\lambda \mathbb{E}[Z^2]}{2(1-\lambda \mathbb{E}[Z])}, \quad (1)$$

where Z is the service time of a single file broadcast. The distribution of Z will depend on the transmission strategy employed at the base station.

Each user will decide to enter if and only if its net utility from the file transfer is non-negative. More specifically, a user will enter the system if and only if

$$U^{(N,K)} - Np - q\mathbb{E}[D(\gamma, p, N)] \geq 0, \quad (2)$$

where $\mathbb{E}[D(\gamma, p, N)]$ is the expected delay experienced by the user, and $q > 0$ is a constant, which we introduce to change the units of delay from time units to monetary units.

This implies that the *effective input rate* λ is given by

$$\begin{aligned} \lambda &= \gamma \mathbb{P}\left(U^{(N,K)} - Np - q\mathbb{E}[D(\gamma, p, N)] \geq 0\right) && (3) \\ &= \gamma \int_{U^{(N,K)} \geq Np + q\mathbb{E}[D]} f_{U^{(N,K)}}(u)\, du, && (4) \end{aligned}$$

where $f_{U^{(N,K)}}(\cdot)$ gives the probability density function of $U^{(N,K)}$. We drop the dependence of D on γ, p and N for ease of exposition. Throughout this paper, we adopt the following assumption on the utility $U^{(N,K)}$.

Assumption 1: $U^{(N,K)}$ is uniformly distributed over the interval $[0, Nb(K)]$, where $b(K)$ is a non-decreasing concave function of the file size K.

The concave dependence of the upper support of the utility value on the file size K suggests that the utility derived from file transfer has diminishing returns. This is a standard assumption in the literature and leads to a tractable analysis.

Under Assumption 1, the relation in (4) simplifies to

$$\frac{1}{Nb(K)}\left(Nb(K) - Np - q\mathbb{E}[D]\right) = \frac{\lambda}{\gamma}. \quad (5)$$

The revenue π generated by the base-station per unit time is the amount each user pays, times the rate at which users enter the system, i.e. $\pi = Np\lambda$. Therefore, the base station's profit maximization problem can be written as

$$\begin{aligned} \max_{\lambda \geq 0, p \geq 0} \quad & Np\lambda, && (6) \\ \text{subject to} \quad & \frac{1}{Nb(K)}(Nb(K) - Np - q\mathbb{E}[D]) = \frac{\lambda}{\gamma}, && (7) \\ & \lambda \leq \frac{1}{X_1}. && (8) \end{aligned}$$

The constraint $\lambda \leq 1/X_1$ is necessary in order for the expected delay to be nonnegative [cf. (1)]. The model we have outlined corresponds to a dynamic game with the following timing of events:

- The base-station sets an entry price p.
- Incoming users decide whether or not to accept the services of the base-station given p.

Characterizing the optimal price p_{opt} and the optimal file size K_{opt} from the perspective of the base-station corresponds to finding the subgame perfect equilibrium of this dynamic two-stage game. Here, every p defines a different subgame. The subgame perfect equilibrium of this game is given by the optimal solution of problem (6) and the corresponding input effective rate λ [cf. (4)]. The above game can also be viewed as a Stackelberg game [3], with the base-station as the leader and potential users as the followers.

III. Revenue Optimization

We next characterize the optimal solution to problem (6). It can be seen that the objective function is continuous and the constraint set is compact, and therefore there exists an optimal solution to (6) denoted by λ_{opt}. Also note that in order to have a finite $\mathbb{E}[D]$ in (7), λ should satisfy $\lambda < 1/X_1$ and thus $\lambda_{opt} < 1/X_1$. This implies that the optimal Lagrange multiplier associated with (8) must be zero due to the slackness constraint. Therefore, it is omitted from the subsequent discussion. In order to find λ_{opt}, we first construct the Lagrangian function $\mathcal{L}(\lambda, p, \mu)$ for problem (6), which is given by

$$\begin{aligned} \mathcal{L}(\lambda, p, \mu) &= Np\lambda \\ &\quad + \mu\left[1 - \frac{q\lambda X_2}{2Nb(K)(1-\lambda X_1)} - \frac{p}{nb(K)} - \frac{\lambda}{\gamma}\right], \end{aligned}$$

where $b(K)$ defines the utility of incoming users (cf. Section 2), μ is the Lagrange multiplier for constraint (7), and X_1 and X_2 are the first and second moments of the service time distribution at the base-station, respectively. Then, the first order optimality conditions for problem (6) yield the following relation between p and λ:

$$p = \lambda\left(\frac{qX_2}{2N(1-\lambda X_1)^2} + \frac{b(K)}{\gamma}\right). \quad (9)$$

Together with the feasibility constraint [cf. (5)], we obtain the following cubic equation in λ:

$$a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0, \quad (10)$$

where

$$\begin{aligned} a_1 &= 4Nb(K)X_1^2, \\ a_2 &= -(8Nb(K)X_1 + 2Nb\gamma X_1^2 + \gamma q X_1 X_2), \\ a_3 &= (4Nb(K) + 2\gamma q X_2 + 4Nb(K)\gamma X_1), \\ a_4 &= -2Nb(K)\gamma. \end{aligned}$$

Since there exists an optimal solution to problem (6), the optimal admission rate λ_{opt} is a solution to the above equation. Our goal is to understand the dependence of λ_{opt} on the number of receivers N and the file size K. In the next section, we show that, when network coding is used, X_1 and X_2 can be expressed as functions of N and K through the use of extreme value theory.

A. Extreme Value Theory

In order to better understand the behavior of λ_{opt} that is described by (10), we need to characterize X_1 and X_2 as functions of N and K. To that end, we use results from Extreme Value Theory, which is stated next.

Theorem 1 ([4]): Let $h_1, \ldots, h_N$ be i.i.d. real random variables with a common distribution function $F(h)$ and density $f(h)$ satisfying the following conditions:

(a) $F(h)$ is twice differentiable for all h.

(b) $f(h)$ is such that

$$\lim_{h\to\infty} \frac{\mathrm{d}}{\mathrm{d}h}\left[\frac{1-F(h)}{f(h)}\right] = 0. \tag{11}$$

Let l_N be such that $F(l_N) = 1 - \frac{1}{N}$. Then, the random variable given by

$$\max_{1\le i\le N} K f(l_N)(h_i - l_N), \tag{12}$$

converges in distribution to a random variable as $N \to \infty$ with cumulative distribution function $\exp(-e^{-x})$, and mean 0.5772, which is the *Euler-Mascheroni* constant.

Let Z denote the completion time of a single file broadcast. It was shown in [6] that when network coding is used, Z is the maximum of N Pascal variables. It is difficult to find exact, closed-form expressions for the first and second moments of Z. The Pascal distribution is a discrete-valued distribution and does not have a continuous, invertible density function. Our current formulation, therefore, does not readily lend itself to extreme value theory. However, a Pascal distribution of order K describing the number of experiments until K successes are achieved can be approximated by an Erlang distribution of order K if the probability of success c in every experiment is sufficiently small [14]. In the following, we adopt this approximation and make the following assumption:

Assumption 2: The mean channel rate $c_n = \frac{\mu_n}{h(N)}$, where $h(N)$ is some monotonically increasing function of N with $\lim_{N\to\infty} h(N) = \infty$, and $\mu_n > 0$ is a constant.

Assumption 2 implies that as the number of receivers N increases, channel conditions between the base-station and the receiver deteriorate, which in turn implies that the probability of a successful packet transmission, i.e. c_n for channel n, becomes smaller. This scenario is particularly relevant in the case where multiple transmitters are situated in the vicinity of the base station, and it is not possible to disregard the possibility of packet erasure due to interference with another transmission. The transmission probability c_n is proportional to the constant μ_n. Therefore, μ_n can be considered to be a measure of the reliability of the channel. In general, the larger μ_n is, the better the chances of a successful transmission are.

We will concentrate on symmetric channel conditions in this paper in order to avoid technical complications, i.e., we will set $\mu_n = \mu > 0$ for all $n \in \{1, \ldots, N\}$. Under Assumption 2, $c_n \to 0$ as $N \to \infty$, and the service time at the base station converges in distribution to a random variable T, which is the maximum of N Erlang variables. The analysis for the general case follows the same line of argument.

Lemma 1: The Erlang distribution satisfies the conditions of Theorem 1.

Proof: The Erlang distribution of order K and rate μ has probability density $f(x)$ and cumulative distribution function $F(x)$ given by

$$\begin{aligned} f(x) &= \frac{\mu^K x^{K-1} e^{-\mu x}}{(K-1)!}, \\ F(x) &= \frac{\gamma(K,\mu x)}{(K-1)!} = 1 - e^{-\mu x}\sum_{i=0}^{K-1}\frac{(\mu x)^i}{i!}, \end{aligned}$$

where $\gamma(\cdot)$ is the *incomplete gamma function*.

It follows that $F(x)$ is twice differentiable for all x. We next show that $f(x)$ and $F(x)$ satisfy (11). We have

$$\begin{aligned} \frac{1-F(x)}{f(x)} &= \frac{e^{-\mu x}\sum_{i=0}^{K-1}\frac{(\mu x)^i}{i!}}{\frac{\mu^K x^{K-1} e^{-\mu x}}{(K-1)!}}, \\ &= \frac{(K-1)!}{\mu^K}\sum_{i=0}^{K-1}\frac{\mu^i}{i!}x^{i-K+1}, \end{aligned}$$

which implies that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1-F(x)}{f(x)}\right] = \frac{(K-1)!}{\mu^K}\sum_{i=0}^{K-1}\frac{\mu^i}{i!}(i-K+1)x^{i-K}.$$

Since $i < K$ for all i in the summation above, each term in the summation goes to zero as $x \to \infty$. Since there are a finite number of terms in the summation, the whole expression goes to zero as $x \to \infty$. Thus

$$\lim_{x\to\infty}\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1-F(x)}{f(x)}\right] = 0,$$

showing that (11) is satisfied, and completing the proof. ■

We use Theorem 1 to characterize the distribution of the completion time of a single file Z. By definition, $Z = \max_{1\le i\le N} Y_i$, where Y_i (the completion time for receiver i) is an Erlang distributed random variable of order K and rate μ, representing the completion time of the file for Receiver-i. If we now use (12) and Theorem 1, a simple linear transformation of variables shows that as $x \to \infty$, Z converges in distribution to a limiting random variable with cumulative distribution function

$$\exp(-e^{-(z-l_N)Nf(l_N)}),$$

with first and second moments given by

$$\begin{aligned} X_1 \equiv \mathbb{E}[Z] &= l_N + \frac{0.5772}{Nf(l_N)}, \\ X_2 \equiv \mathbb{E}[Z^2] &= (\mathbb{E}[Z])^2 + \frac{1}{6}\frac{\pi^2}{(Nf(l_N))^2}. \end{aligned}$$

In order to completely characterize the first and second moments of Z, we need an explicit expression for l_N.

Theorem 1 defines l_N implicitly as $F(l_N) = 1 - \frac{1}{N}$. Replacing $F(.)$ by the expression for the cumulative distribution function of the order K-Erlang, we obtain

$$\begin{aligned} F(l_N) &= 1 - \frac{1}{N}, \\ 1 - e^{-\mu l_N} \sum_{i=0}^{K-1} \frac{(\mu l_N)^i}{i!} &= 1 - \frac{1}{N}, \\ e^{-\mu l_N} \sum_{i=0}^{K-1} \frac{(\mu l_N)^i}{i!} &= \frac{1}{N}. \end{aligned}$$

We assume that $\mu l_N \gg K$, which is a reasonable assumption since l_N diverges to infinity as N increases. Thus, for a sufficiently large number of receivers, this assumption holds. Then, the last term in the summation would dominate, and we can omit all the other terms and retain the last one. Therefore, the above summation simplifies to:

$$\frac{\mu^K l_N^{K-1} e^{-\mu l_N}}{(K-1)!} = \frac{1}{N}.$$

Even though this expression is still intractable, we can use it to obtain upper and lower bounds on μl_N. If we substitute $\mu l_N = \log N$ in the expression, we obtain

$$\frac{\mu}{(K-1)!} \frac{(\log N)^{(K-1)}}{N},$$

which is $\Omega(1/N)$. Therefore, μl_N can be lower-bounded by $\log N$. Furthermore, if we substitute $\mu l_N = c \log N$, where c is a positive constant greater than 1, we obtain

$$\frac{\mu c^{(K-1)}}{(K-1)!} \frac{(\log N)^{(K-1)}}{N^c},$$

which, for a sufficiently large c, is $O(1/N)$. Therefore, μl_N can be upper-bounded by $c \log N$. In other words, there exists an N_0 such that for all $N > N_0$, $\log N \leq \mu l_N \leq c \log N$. More precisely, l_N behaves as $\Theta(logN)$ as $N \to \infty$. Given the asymptotic behavior of l_N, upper and lower bounds can be obtained on the first and second moments of Z.

If we plot X_1 and X_2 as a function of the file size K, then, for N sufficiently large, we observe that X_1 is almost linear in K for a large range of values of K. In fact, X_1 and X_2 can be approximated as:

$$X_1 = \frac{1.15(K-1) + 2.42log(N)}{\mu}, \qquad X_2 = X_1^2.$$

These approximations are plotted for $N = 50$ and $\mu = 1$ in Figures 2 and 3 along with the actual X_1 and X_2.

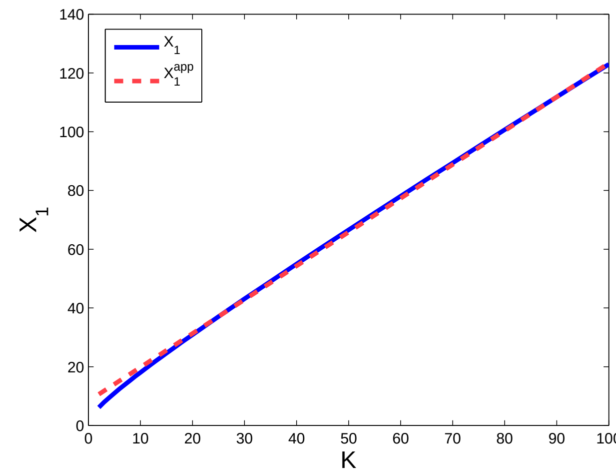

Fig. 2. Approximation of X_1 as a function of file size K, $N = 50$, $\mu = 1$

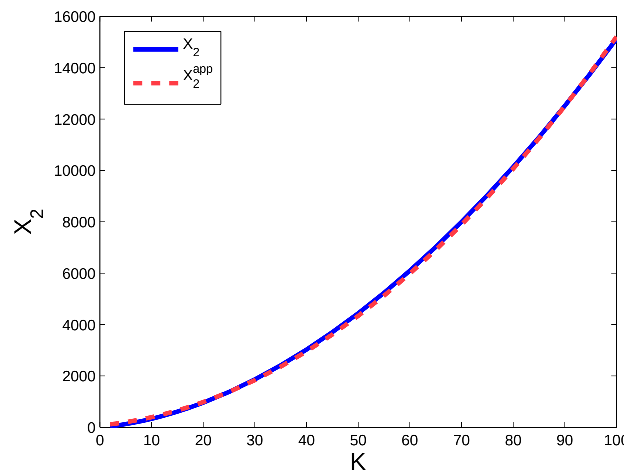

Fig. 3. Approximation of X_2 as a function of file size K, $N = 50$, $\mu = 1$

B. Optimal admission rate as a function of system parameters

Among the solutions of (10), we pick the one that yields the maximum revenue and is a feasible solution of (6).

We adopt the following assumption in our analysis:

Assumption 3: The number of receivers N is sufficiently large, and the arrival rate γ is sufficiently small so that $\gamma q X_2 \ll 2b(K)N$.

With this assumption, we focus our attention on the low-traffic regime with relatively dense network model. Note that this does not imply low throughput since the number of receivers is high and hence the aggregate throughput will typically be large. Also note that it is worthwhile to investigate the behavior of the optimal admission rate in the low-traffic mode, since in reality downlink systems are designed and deployed to avoid an inordinately large number of users demanding service, and the assumption that the base-station is operating in the low-traffic regime is a realistic one.

An inspection of the coefficients in (10) shows that all the terms are multiples of N, except for the terms $\gamma q X_1 X_2$ in a_2 and $2\gamma q X_2$ in a_3. Interestingly, these two terms are also multiples of q. No other term contains q. Consequently, if q is small, then by using Assumption 3, we can modify the terms $\gamma q X_1 X_2$ in a_2 and $2\gamma q X_2$ in a_3 without significantly affecting the values of the coefficients a_2 and a_3. Our goal will be to use the roots of the modified cubic equation to obtain simpler expressions for the roots of the original cubic equation. If we change the term $2\gamma q X_2$ in a_3 to $\gamma q X_2$, without altering any other term, the resulting cubic equation admits the following three roots, which under Assumption 3 are close approximations to the roots of the original cubic equation:

$$\begin{aligned}\lambda_1 &= \frac{1}{X_1},\\ \lambda_2 &= \frac{1}{2X_1} + \frac{\gamma}{4} + \frac{\gamma q X_2}{8b(K)NX_1} + \psi\\ \lambda_3 &= \frac{1}{2X_1} + \frac{\gamma}{4} + \frac{\gamma q X_2}{8b(K)NX_1} - \psi,\end{aligned}$$

where

$$\psi \triangleq \frac{\sqrt{(2b(K)N\gamma X_1 - 4b(K)N + \gamma q X_2)^2 + 16b(K)N\gamma q X_2}}{8b(K)NX_1}.$$

Out of these three roots, only $\lambda_3 < 1/X_1$ and is therefore the optimal solution of (6). In order to gain more insight into the expression for the optimal admission rate, we rewrite λ_3 as follows:

$$\lambda_3 = \frac{y - \sqrt{y^2 - z}}{8b(K)NX_1},$$

where $y \triangleq 4b(K)N + 2b(K)\gamma NX_1 + \gamma q X_2$, and $z \triangleq 32b(K)^2\gamma N^2 X_1$. The term within the square root can be expressed as
$\sqrt{y^2 - z}$

$$= y\sqrt{1 + \frac{(2+\gamma X_1)}{(2-\gamma X_1)^2}\frac{\gamma q X_2}{b(K)N} + (\frac{\gamma q X_2}{4b(K)N - 2b(K)\gamma NX_1})^2}$$

By Assumption 3, we can neglect the term in $(\frac{\gamma q X_2}{b(K)N})^2$. To simplify the expression for the optimal effective input rate, we further adopt the following assumption:

Assumption 4: If Assumption 3 holds, then $\frac{(2+\gamma X_1)}{(2-\gamma X_1)^2}\frac{\gamma q X_2}{b(K)N} \ll 1$.

Proposition 1: Let Assumptions 3 and 4 hold. Then the optimal effective input rate is given by

$$\lambda_{opt} = \lambda_3 = \begin{cases} \frac{\gamma}{2} - \frac{(\gamma)^2 q X_2}{4b(K)N(2-\gamma X_1)} & \gamma X_1 < 2 \\ \frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)} & \gamma X_1 > 2 \end{cases}$$

Proof: Use a first order Taylor approximation to write

$$\sqrt{1 + \frac{(2+\gamma X_1)}{(2-\gamma X_1)^2}\frac{\gamma q X_2}{b(K)N}} \approx 1 + \frac{1}{2}\frac{2+\gamma X_1}{(2-\gamma X_1)^2}\frac{\gamma q X_2}{b(K)N},$$

Under Assumptions 3,4, λ_3 can be written as

$$\frac{y - \sqrt{(4b(K)N - 2b(K)\gamma NX_1)^2}\left(1 + \frac{1}{2}\frac{2+\gamma X_1}{(2-\gamma X_1)^2}\frac{\gamma q X_2}{b(K)N}\right)}{8b(K)NX_1}.$$

We note that the term $(4b(K)N - 2b(K)\gamma NX_1)$ is positive for $\gamma X_1 < 2$, and is negative for $\gamma X_1 > 2$. Therefore, we consider the following two cases, and derive a piecewise expression for λ_3.

CASE 1: $\gamma X_1 < 2$.

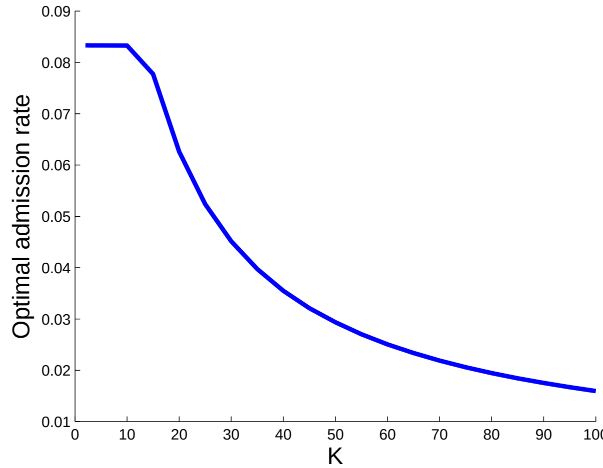

Fig. 4. Plot of optimal admission rate as a function of file size K, $N = 50$, $\mu = 2$

In this situation, $(4b(K)N - 2b(K)\gamma NX_1)$ is positive, and therefore

$$\begin{aligned}\lambda_3 &= \frac{4b(K)N\gamma X_1 + \gamma q X_2 - \gamma q X_2\left(\frac{2+\gamma X_1}{2-\gamma X_1}\right)}{8b(K)NX_1}\\ &= \frac{\gamma}{2} - \frac{(\gamma)^2 q X_2}{4b(K)N(2-\gamma X_1)}.\end{aligned}$$

CASE 2: $\gamma X_1 > 2$.

In this case, $(4b(K)N - 2b(K)\gamma NX_1)$ is negative, and we must negate it in order to take the positive square root. Therefore

$$\begin{aligned}\lambda_3 &= \frac{8b(K)N + \gamma q X_2 - \gamma q X_2\left(\frac{2+\gamma X_1}{\gamma X_1 - 2}\right)}{8b(K)NX_1}\\ &= \frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)}.\end{aligned}$$

This completes the proof. ■

Note that, λ_{opt} is constant at $\gamma/2$ for small values of K, i.e. for $\gamma X_1 < 2$, and decreases approximately as $1/X_1$ for larger values of K, i.e. $\gamma X_1 > 2$. Since the constraint $\gamma X_1 > 2$ is the same as $\gamma/2 > 1/X_1$, we can write

$$\lambda_{opt} = \begin{cases} \frac{\gamma}{2} - \frac{(\gamma)^2 q X_2}{4b(K)N(2-\gamma X_1)} & \gamma/2 < 1/X_1 \\ \frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)} & \gamma/2 > 1/X_1 \end{cases}$$

which can essentially be re-written as

$$\lambda_{opt} = \min\{\gamma/2, 1/X_1\} - f(\gamma, N, X_2),$$

for appropriately defined $f(\cdot)$.

Figure 4 shows the optimal admission rate as a function of the file size K for $N = 50$ and $\mu = 2$.

IV. Optimal Price and Revenue

In this section, we study the optimal price and revenue behavior of our system with changing system parameters. Figure 5 plots the optimal revenue of the base station as a function of the corresponding file size K for varying values of N. We observe two features:

- The optimal revenue is a unimodal function of K, with a single stationary point
- The file size K_{opt} which maximizes the revenue does not change as the number of receivers N is varied.

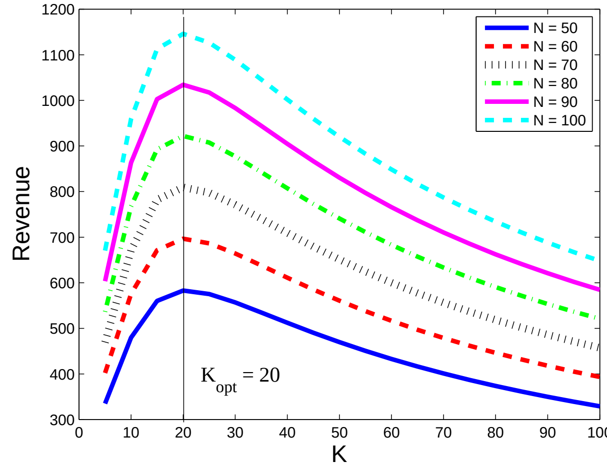

Fig. 5. Plot of revenue as a function of K for various values of N, with $\mu = 2$.

We next characterize the optimal price and revenue under Assumptions 3 and 4.

Proposition 2: Let Assumptions 3 and 4 hold. Let p_{opt} be the price that maximizes problem (6) and π_{opt} be the corresponding optimal profit. Then,

$$p_{opt} = \begin{cases} \frac{b(K)}{2} + \frac{\gamma q X_2}{4N(1-\frac{\gamma X_1}{2})^2}, & \gamma X_1 < 2 \\ \frac{1}{X_1}\left[\frac{2Nb(K)^2(\gamma X_1 -2)^2}{(\gamma)^2 q X_2} + \frac{b(K)}{\gamma} \right. \\ \left. -\frac{b(K)(\gamma X_1 - 2)}{\gamma} - qX_2 2N(\gamma X_1 - 2)\right], & \gamma X_1 > 2 \end{cases}$$

$$\pi_{opt} = \begin{cases} \frac{(\gamma)^2 q X_2}{2(2-\gamma X_1)^2} + \frac{Nb(K)\gamma}{4}, & \gamma X_1 < 2 \\ \frac{1}{X_1^2}\left[\frac{2N^2 b(K)^2(\gamma X_1 - 2)^2}{(\gamma)^2 q X_2} + \frac{Nb(K)}{\gamma}\right], & \gamma X_1 > 2 \end{cases}$$

Proof:

CASE 1: If $\gamma X_1 < 2$, then under Assumption 3, $\frac{(\gamma)^2 q X_2}{4b(K)N(2-\gamma X_1)}$ is negligible compared to $\gamma/2$ and $\lambda_{opt} \approx \gamma/2$. Therefore

$$\begin{aligned} p_{opt} &= \lambda_{opt}\left[\frac{qX_2}{2N(1-\frac{\gamma X_1}{2})^2} + \frac{b(K)}{\gamma}\right] \\ &= \frac{b(K)}{2} + \frac{\gamma q X_2}{4N(1-\frac{\gamma X_1}{2})^2} \\ \pi_{opt} &= N p_{opt} \lambda_{opt} \\ &= (\lambda_{opt})^2\left[\frac{qX_2}{2(1-\frac{\gamma X_1}{2})^2} + N\frac{b(K)}{\gamma}\right] \\ &= \frac{b(K)}{2} + \frac{\gamma q X_2}{4N(1-\frac{\gamma X_1}{2})^2} \\ &= \frac{(\gamma)^2 q X_2}{2(2-\gamma X_1)^2} + \frac{Nb(K)\gamma}{4}. \end{aligned}$$

We notice that when the file size K is small enough so that $\gamma X_1 < 2$, then under Assumption 3, $p_{opt} \approx b(K)/2$, and $\pi_{opt} \approx Nb(K)\gamma/4$.

CASE 2: If $\gamma X_1 > 2$, then $\lambda_{opt} = \frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)}$. Therefore,

$$\begin{aligned} p_{opt} &= \lambda_{opt}\left[\frac{qX_2}{2N(1-\frac{\gamma X_1}{2})^2} + \frac{b(K)}{\gamma}\right] \\ &= \left(\frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)}\right) \times \\ &\quad \left[\frac{qX_2}{2N\left(\frac{\gamma q X_2}{2b(K)N(\gamma X_1 - 2)}\right)^2} + \frac{b(K)}{\gamma}\right] \\ &= \frac{1}{X_1}\left[\frac{2Nb(K)^2(\gamma X_1 - 2)^2}{(\gamma)^2 q X_2} + \frac{b(K)}{\gamma} \right. \\ &\quad \left. -\frac{b(K)(\gamma X_1 - 2)}{\gamma} - qX_2 2N(\gamma X_1 - 2)\right]. \end{aligned}$$

$$\begin{aligned} \pi_{opt} &= (\lambda_{opt})^2\left[\frac{qX_2}{2(1-\frac{\gamma X_1}{2})^2} + N\frac{b(K)}{\gamma}\right] \\ &= \left(\frac{1}{X_1} - \frac{\gamma q X_2}{2b(K)NX_1(\gamma X_1 - 2)}\right) \times \\ &\quad \left[\frac{qX_2}{2\left(\frac{\gamma q X_2}{2b(K)N(\gamma X_1 - 2)}\right)^2} + \frac{Nb(K)}{\gamma}\right] \\ &= \frac{1}{X_1^2}\left(2b(K)N(\gamma X_1 - 2)^2 - \gamma q X_2\right) \times \\ &\quad \left[\frac{1}{2(\gamma)^2 q X_2} + \frac{1}{4\gamma b(K)N(\gamma X_1 - 2)^2}\right] \\ &\approx \frac{1}{X_1^2}\left[\frac{2N^2 b(K)^2(\gamma X_1 - 2)^2}{(\gamma)^2 q X_2} + \frac{Nb(K)}{\gamma}\right]. \end{aligned}$$

■

These expressions are only approximate expressions because the revenue and price functions are extremely sensitive to λ. This is due to the presence of the $(1 - \lambda X_1)$ term in the denominator and the fact that λ_{opt} is very close to $1/X_1$.

1) *The optimal revenue is a unimodal function of the file size:* We note that for $\gamma X_1 < 2$, the dependence of

optimal revenue on K is given by $Nb(K)\gamma/4$. For $\gamma X_1 > 2$, the revenue function contains terms in $b(K)/X_1^2$ and $b(K)^2/X_1^4$. Since $b(K)$ is a concave function in K, and X_1 is linear in K (cf. Section III-A), the revenue function is monotonically nonincreasing. We note further that since the revenue is monotonically increasing in the region $\gamma X_1 < 2$, the optimum file size K_{opt} which maximizes the revenue occurs in the range $\gamma X_1 > 2$.

2) *The file size that maximizes the revenue is insensitive to changes in N:* We next find the file size that maximizes the revenue, denoted by K_{opt} (cf. Figure 5). We know that K_{opt} occurs in the range $\gamma X_1 > 2$, so we need only look at the revenue function in that range. We assume $b(K) = c\log(K)$ where c is a constant. Using the approximations for X_1 and X_2 from Section III-A and taking first derivatives with respect to K, we obtain the following implicit expression in K and N, which characterizes K_{opt} in terms of N.

$$2\gamma^2\left(2X_1^3\frac{log(K)}{k} - 2.3X_1^2 log(K)^2\right)$$
$$-4\gamma\left(2X_1^2\frac{log(K)}{K} - 3.45X_1 log(K)^2\right)$$
$$+8\left(2X_1\frac{log(K)}{K} - 4.6 log(K)^2\right)$$
$$-\frac{\gamma q}{Nc}\left(\frac{X_1^3}{K} - 2.3X_1^2 log(K)\right) = 0$$

The above equation is a transcendental equation, and does not admit a tractable solution. However, our primary objective is not to solve for K_{opt}, but to understand its relative stability as the number of receivers N changes. Towards this end, we differentiate the above equation implicitly with respect to N in order to obtain an expression for $\partial K_{opt}/\partial N$.

From Figure 5, we observe that $K_{opt} \approx 20$. We can calculate the values of $\partial K_{opt}/\partial N$ for $K_{opt} \approx 20$ and different values of N. If we let $N = 50$, the value of $\partial K_{opt}/\partial N$ turns out to be approximately 0.003. In other words, if N changes by 100, K_{opt} would change by only 0.3. In fact, for the range of N that is of interest to us, K_{opt} is constant as shown in Figure 5.

In the above analysis, we assumed that $b(K) = a\log K$. However, it can be shown that the same implications hold for other concave functions such as $\sqrt{K}$ and functions of the form $K^{1/p}$ where $p > 1$.

V. Sensitivity to System Parameters

We next study through simulations the dependence of the optimal admission rate, the optimal price and the expected delay as functions of the file size K. Figures 4, 5 6 and 7 show plots of the optimal admission rate, the optimal revenue, the optimal price, and the expected delay respectively as functions of K.

The optimal admission rate offers a great deal of insight into the dynamics of the system. As noted in Section III-B, the optimal admission rate is approximately $\gamma/2$ for small values of K, and approximately $1/X_1$ for larger values of K. Therefore, if the system is operating in the low-traffic regime, and the file size is small, the optimal admission rate is $\gamma/2$. Intuitively, this makes sense because the queueing delay is negligible (Figure 7), and the major cost experienced by the users is price. As noted earlier (cf. Section IV), price per receiver increases as $b(K)/2$ for small values of K, and given that the utility function is uniform between $[0, Nb(K)]$ (cf. Assumption 1), we would expect half of the arrivals to accept the service, and half to reject it.

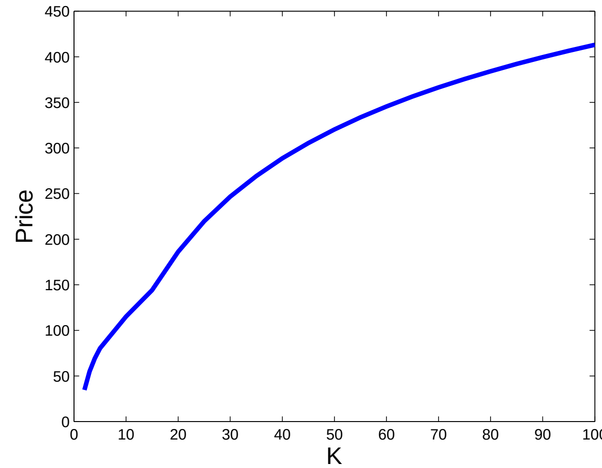

Fig. 6. Plot of price as a function of file size K, $N = 50$, $\mu = 2$

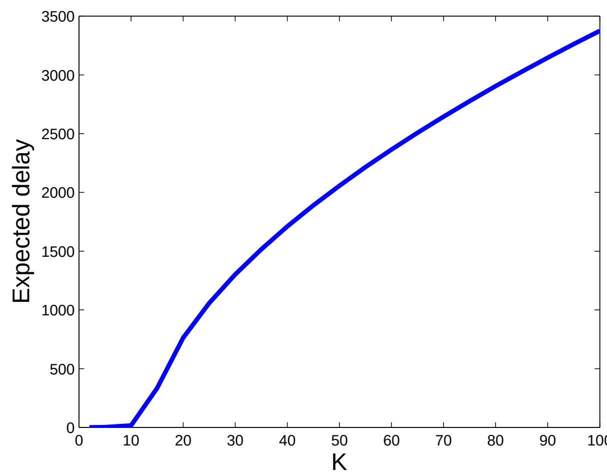

Fig. 7. Plot of expected delay as a function of file size K, $N = 50$, $\mu = 2$

As the file size progressively increases, the expected queueing delay also increases. The threshold at which the effects of delay can no longer be ignored is $\frac{\gamma}{2} = \frac{1}{X_1}$. As this threshold is crossed, the effects of the delay become sufficiently appreciable, and the expected service time X_1 increases rapidly. The optimal admission rate is then constrained by the reciprocal of the mean service time. The second moment X_2 plays an important role in this case. It perturbs the optimal admission rate so that it is slightly below

$1/X_1$. If X_2 were zero (physically impossible since this would mean that the service time has a negative variance), then λ_{opt} would be exactly $1/X_1$, which would lead to an infinite revenue, infinite price, and an infinite delay (all of which are physically impossible). The effect of X_2 implies that the users' decision is also affected by the variance of the service time. A higher variance will lead to a lower admission rate.

The optimal revenue is proportional to the product of the optimal price and the optimal admission rate. For smaller values of K, the revenue of the base-station increases because the queuing delay is not very significant and the admission rate is constant. The base-station can therefore increase its price with the guarantee that the admission rate will not decrease as long as the file size is small. For larger file sizes, the delay becomes significant and the base-station is no longer able to increase its price. The optimal admission rate begins to decrease as a result of large queuing delays. Consequently, the revenue of the base-station reaches a maximum, and then begins to decrease as K increases. Qualitatively, this describes the unimodal shape of the optimal revenue function depicted in Figure 5.

The expected delay increases rapidly (Figure 7), in fact almost linearly, as K increases. This increase in the expected delay is due to the fact that as the file size increases, it takes the base-station progressively longer to transmit the file to the receivers, which in turn increases the waiting time for other users in the queue.

The optimal price (Figure 6) initially increases as K increases, but eventually tapers off in a sub-linear fashion. This observation can be understood if we think of the price and the expected delay as two different costs that users will experience upon entering the system. As the expected delay increases and the optimal admission rate begins to drop, the base-station cannot afford to keep increasing its price, since that would exacerbate the drop in the optimal admission rate. In order to mitigate the effect of the increased delay on the admission rate, the base-station must check its price in order to encourage more users to join the system.

The dynamics of the system are also affected by the channel conditions. The quality of the channel is captured by μ (cf. Assumption 1). From the analysis of Section III-A, we know that X_1 is inversely proportional to μ. Therefore, if μ is larger (i.e. the channel conditions are better), the mean service time will be smaller, and files will be transferred more quickly. Since X_1 will be smaller, the threshold $\gamma/2 = 1/X_1$ will be crossed at a much larger value of K. In other words, the optimal admission rate will remain constant at $\gamma/2$ for larger file sizes as well. The base-station will be able to increase its revenue over a larger range of values of K. The optimum file size K_{opt} will also be larger. This result is shown in Figure 8. The optimum file size increases almost linearly with μ. The implication, therefore, is that given better channel conditions, the base station will choose to transfer larger files, and will attain higher profits at the same time.

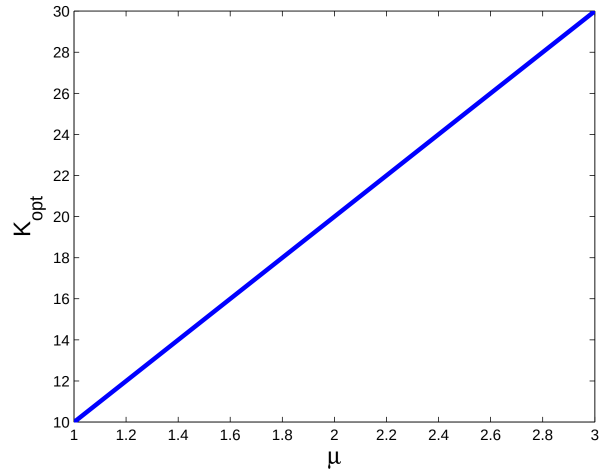

Fig. 8. Plot of K_{opt} as a function of μ, $N = 50$

VI. Coding vs. Scheduling

In order to quantify the economic gains from network coding, we must compare it with the traditional scheduling approach, which is currently the common mode of packet transmission. A comprehensive analysis of various scheduling policies with and without CSI is given in [6]. The authors find that the optimal scheduling policy for a system with no CSI is the *round-robin* approach, in which the base station sends a single packet to every receiver in turn. We denote the first and second moments of the service time distribution in the round-robin case by X_1^{RR} and X_2^{RR}, and the first and second moments of the service time in the network coding case by X_1^{NC} and X_2^{NC}.

It is shown in [6] that X_1^{RR} is upper-bounded as

$$X_1^{RR} \leq K + K\mathbb{E}[\max_{1\leq k\leq K, 1\leq n\leq N} U_n^k],$$

where U_n^k is a geometric random variable with parameter c, representing the number of slots until a given channel is ON. The geometric distribution converges to exponential distribution with rate μ as N gets large. Again, we use extreme value theory to compute the upper bound, and then use the fact that $X_2^{RR} \geq (X_1^{RR})^2$ in order to get a lower bound on X_2^{RR}. Since a larger value of X_1^{RR} and a lower value of X_2^{RR} improves the performance, we use the above bounds to compute an upper bound on the performance of round-robin. In the following we use this upper bound for comparison of the two transmission strategies.

Since the mean service time of round-robin scheduler is larger [6], we would expect the threshold $\gamma/2 = 1/X_1^{RR}$ to be crossed at a much smaller file size than that for network coding, and the revenue of the base station would start to decrease for small K. In other words, K_{opt} for round-robin will be very small, and thus the revenue earned by the base station will be much lower compared to the revenue from network coding. Figures 9 and 10 present a comparison of the optimal admission rate and revenue from network coding and round-robin, respectively. As expected, the revenue in the round-robin case begins to decrease at a much smaller value of K, and the difference between the revenues from network coding and round-robin is also very significant. The optimal admission rate for round-robin is also much lower than that for network coding, while the expected delay in the case of round-robin is considerably larger.

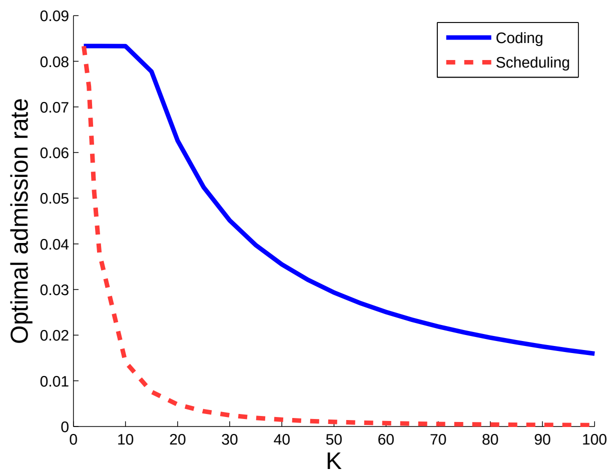

Fig. 9. Comparison of optimal admission rate with network coding and round-robin as a function of file size K , $N = 50$ $\mu = 2$

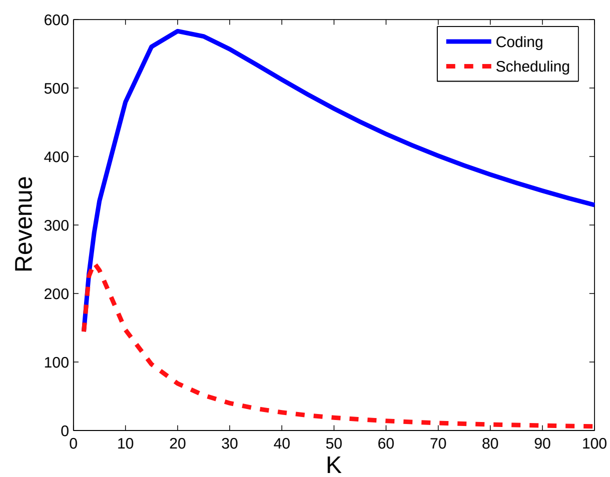

Fig. 10. Comparison of revenue with network coding and round-robin against file size K , $N = 50$ $\mu = 2$

VII. Conclusions

In this work, we have looked at a simple file transfer system with a single base station and multiple receivers. We have investigated the system from an economic viewpoint, and have obtained characterizations of the optimal user admission rate and its dependence on the moments of the service time at the base station. We have also approximated the optimal price and the maximum revenue charged by the base station, and have shown that the revenue is a unimodal function of the file size with a single maximum. Furthermore, the optimum file size,i.e., the file size for which the revenue is maximized, is highly insensitive to changes in the number of receivers. We have also compared the revenue, price and delay obtained from network coding to that obtained from scheduling, and have observed that network coding yields significant gains in revenue while allowing for a higher user admission rate at the same time. We have also discussed the effect of channel conditions on the maximum revenue generated by the base station, and the optimum file size.

The possibilities for further work are vast. We have assumed that the system is operating in low-traffic. An interesting extension would be to characterize the optimal revenue, price and delay in a high-traffic regime, i.e., when the user arrival rate is large. Another direction would be to analyze the system using utility functions other than the uniform distribution.

References

[1] D. Acemoglu, A. Ozdaglar, and R. Srikant. The marginal user principle for resource allocation in wireless networks. In *Proceedings of IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

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

[3] T. Basar and G. J. Oldser. *Dynamic Noncooperative Game Theory*. SIAM Series in Classics in Applied Mathematics, 1999.

[4] H. A. David. *Order Statistics*. John Wiley, Second Edition, 2003.

[5] 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 *Proceedings of International Workshop on Wireless Ad-hoc Networks (IWWAN)*, 2005.

[6] A. Eryilmaz, A. Ozdaglar, and M. Médard. On delay performance gains from network coding. In *Proceedings of Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2006.

[7] T. Ho, M. Médard, and R. Koetter. An information theoretic view of network management. In *Proceedings of Infocom*, 2003.

[8] D. Katabi, S. Katti, W. Hu, H. Rahul, and M. Médard. The importance of being opportunistic: On practical network coding for wireless environments. In *Proceedings of Annual Allerton Conference on Communication, Control and Computing*, October 2005.

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

[10] S.-Y. R. Li, R. W. Yeung, and Ning Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49:371–381, February 2003.

[11] D. S. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee. Achieving minimum cost multicast: A decentralized approach based on network coding. In *Proceedings of IEEE Infocom*, pages 1607–1617, 2005.

[12] C. Magalaras and A. Zeevi. Pricing and performance analysis for a system with differentiated services and customer choice. In *Proceedings of the Allerton Conference on Control, Communications and Computing*, 2003.

[13] I. Ch. Paschalidis and J. N. Tsitsiklis. Congestion-dependent pricing of network services. *IEEE/ACM Transactions on Networking*, 8(2):171–184, 2000.

[14] J. Tsitsiklis and D. Bertsekas. *Introduction to Probability*. Athena Scientific, Belmont, MA, 2002.

[15] Y. Zhu, B. Li, and J. Guo. Multicast with network coding in application-layer overlay networks. *IEEE JSAC*, pages 107–120, January 2004.