Performance Trade-offs in a Network Coding Router

Omer H. Abdelrahman and Erol Gelenbe, *Fellow, IEEE*
Dept. of Electrical & Electronic Engineering
Imperial College London, UK
Email:{o.abd06, e.gelenbe}@imperial.ac.uk

***Abstract*—We consider the problem of optimizing the performance of a network coding router with two stochastic flows. We develop a queueing model which accounts for the fact that coding is not performed when packets are transmitted, but is done by a separate program or hardware which operates independently of the hardware that sends packets out over links. We formulate and solve a constrained optimization problem which provides the optimal time that the router should wait before sending the information that it has uncoded, so that the average response time of the system is minimized. The trade-offs between delay and bandwidth or energy associated with the choice of the waiting time are also investigated, and the results indicate that network coding offers significant performance gains in a moderate to heavily loaded system.**

I. Introduction

In the emerging field of network coding [1], routers are allowed to process and mix information within packets before forwarding them towards their destinations. This approach has the potential of increasing throughput and improving robustness of communication networks. However, its impact on packet delay is not yet fully understood. Indeed, although a lower traffic rate per link will necessarily reduce the link delay, and thus the overall delay that a given packet travelling through a network will experience, network coding can also increase delay in several ways. The need for combining packets at nodes may force packets to wait for the arrival of other packets with which they will be combined, introducing a potential *synchronization* delay. Also, although individual link delays will be reduced, node delays may in fact not be affected because in order to reconstitute the packet streams at output nodes, the nodes will have to carry on the average the same amount of traffic if no information is to be lost, so that congestion would not be reduced or may even be increased by network coding. Finally, the need to decode packets at output nodes implies a further delay for the "right" combination of packets to arrive before a given packet can be decoded and forwarded to the receiver.

The trade-off in network coding between delay and transmission costs (bandwidth and energy) under stochastic packet arrivals has been considered previously. In [2], the energy delay trade-off for a two-way relay network is analyzed assuming that the relay accumulates packets from one direction and sends them either after packets from the other direction arrive or the number of packets waiting exceeds the buffer capacity. Packet transmission is then assumed to occur instantaneously. The analysis indicates that in the case of even traffic load, the average delay must tend to infinity in order to achieve minimum energy consumption. A discrete time analysis of this scenario under probabilistic network coding is presented in [3]. The two-way relay network has also been studied for slotted Aloha [4], and it was shown that the delay throughput trade-off depends on the transmission probability of the relay.

In [5], the stability and energy consumption of network coding in a wireless tandem network with slotted transmission are considered; intermediate nodes can either transmit self-generated packets or encode two relay flows received from neighboring nodes. The results obtained suggest that immediate transmission of first available packets yields higher throughput as compared to waiting for additional packets to arrive before coding, but this gain comes at the expense of reducing energy efficiency. Conversely, we show that, under stochastic arrivals, immediate transmission of packets cannot offer the opportunity for coding.

In [6], the multicast delay and throughput trade-off with intra-flow coding is considered for a slotted-time collision-based wireless network, showing that coding improves throughput and energy costs at the expense of higher packet delays as compared to plain routing. However, the results were obtained under the unrealistic assumptions of one-bit packet lengths and saturated queues at source and relay nodes.

There has been considerable work on evaluating the performance of network coding in different mathematical frameworks. In [7], the achievable rate regions under quality of service (QoS) constraints are computed for a butterfly network with and without network coding. However, the analysis is based on a fluid flow model which does not capture the bursty nature of packet arrivals which is essential for understanding network coding gains. End-to-end QoS bounds for both network coding and plain forwarding have been derived in [8] using deterministic network calculus; the results show that coding can improve the worst case delays even in topologies where no throughput gains are expected. Network calculus, however, can only provide bounds that may not be tight in practice.

The contributions of the present paper are twofold. First, we decouple the different stages of service in a network coding router in order to address the fact that coding is not performed when packets are transmitted, but is performed by a separate program or hardware which operates independently of the hardware that forwards packets. In contrast, existing theoretical work on network coding has assumed either zero transmission time [2], slotted time [3]–[5] or packet length based service time [9]–[11]. Second, our model gives rise

978-1-4244-7115-7/10/$26.00 ©2010 IEEE

to a trade-off study and an optimization problem that have not been considered before. In particular, network coding allows better utilization of network resources but the delay for packet encoding may degrade performance. Reducing this delay, however, will lessen coding opportunities and increase congestion which may subsequently increase delay. This paper investigates the trade-offs associated with the length of the coding delay.

II. System Description

In order to illustrate these trade-offs, we consider a two-way relay network in which two nodes A and B wish to communicate with each other through an intermediate node R, as shown in Fig. 1(a). In the figure, A sends a packet P_1 and B sends P_2 to R which then broadcasts $P_1 \oplus P_2$ instead of the two packets in sequence. Both A and B can extract their desired packets by simply xor-ing the received packet with the transmitted one, therefore reducing the number of required transmissions from 4 to 3. While no synchronization delay is incurred at the decoding nodes in this scenario, this is not necessarily the case at the relay node since it may need to delay packets from one direction until packets from the other direction arrive.

We assume that each source generates packets independently according to a Poisson process with rate $\lambda_i,\ i = 1, 2$ so that the total incoming traffic rate at the intermediate node is $\Lambda = \lambda_1 + \lambda_2$. Denote by $\rho_i = \lambda_i/\Lambda$ the probability that a received packet at the relay node is of class i. The relay's buffers are assumed to be of unlimited capacity so that loss of packets due to buffer overflow cannot occur. The queueing network model for the encoding node is depicted in Fig. 1(b).

When a packet arrives at the relay node, it is stored into an input buffer where its header is processed, then it is forwarded to buffer $i = 1, 2$ of the coding component based on its class. We assume that the service time at the input queue is exponential with parameter δ. The policy employed by the coding component to process the packets from the two buffers is as follows:

- When a packet reaches the head of buffer i, if the other buffer $j = 3 - i$ is not empty, then this packet is immediately xor-ed with the head of the line packet from buffer j and the coded version is placed in the transmission queue. We assume that the xor operation is instantaneous so that no further delay is incurred after the two packets are matched. This is a fairly reasonable assumption given the current processing power.
- When a packet arrives at the head of buffer i and the buffer holding the traffic from class $j = 3 - i$ is empty, then the packet waits for an exponential time-out period with parameter γ_i. If a packet arrives at queue j before the timer elapses, then the two packets will be coded together and the coded version will be placed in the transmission queue. On the other hand, if the timer expires before a packet arrives at buffer j, then the head of the line packet from buffer i will be forwarded to the transmission queue without coding.

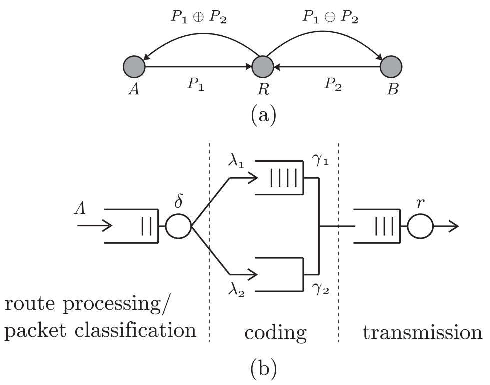


Fig. 1. (a) A two-way relay network. (b) The proposed queueing model for the encoding node R illustrating the different stages of service.

The random transmission time of packets is assumed to follow a distribution with mean $1/r$. In Section III we will assume that this distribution is exponential in order to attain an exact solution. A more realistic model is then analyzed approximately in Section IV when the transmission time is directly proportional to packet length and the latter is constant. In both cases, we assume that network coding does not incur additional overhead apart from the synchronization delay at the coding stage.

Although delaying packets for coding may appear harmful particularly for stochastic arrivals, it turns out that there is a compromise to be achieved between increasing delay at the coding queues on the one hand, and reducing delay at the transmission queue on the other, by tuning the time-out parameters in order to adjust coding opportunities. The question then is how long should a coding node wait before going ahead and sending the information that it has uncoded? Short waiting times imply few coding opportunities and therefore inefficient network operation while long waiting times imply added delay which eventually overtakes the original benefit of coding.

III. Generating Function Analysis

We derive the generating function for the joint coding and transmission queue lengths when arrival rates are symmetric, i.e. $\lambda_i = \lambda$, and transmission time is exponentially distributed. In this case and due to the symmetry of arrivals, the time-out parameters must be chosen to be equal, i.e. $\gamma_i = \gamma$, in order to yield similar performance for both packet classes. Since the route processing stage is independent of the time-out parameters and is identical for both network coding and plain forwarding, in the rest of the paper we will focus on the coding and transmission components of the node.

Denote by $p(m, n)$ the stationary probability of having m packets in the coding queues and n packets in the transmission queue, i.e. $p(m, n) = \lim_{t\to\infty} P[Q_c(t) = m, Q_{tr}(t) = n]$. Note that due to the coding policy described previously, at most one of the two coding queues can be non-empty. Moreover, because of the symmetry of arrivals and time-outs, the stationary probability that there are m packets in the ith coding queue is related to the stationary distribution of the aggregate number of packets in both queues by $P[Q_c^i = m] = \frac{1}{2}P[Q_c = m]$, for $m > 0$. Hence, it is sufficient to analyze the coding stage as a single queue.

The balance equations for the probabilities $p(m,n)$ satisfy:

$$\begin{aligned}\Big[(\lambda+\mu)1_{\{m>0\}} + 2\lambda 1_{\{m=0\}} + r1_{\{n>0\}}\Big]&p(m,n)\\ = \; rp(m,n+1) + \mu p(m+1,n-1)&1_{\{n>0\}}\\ + \; \big[\lambda 1_{\{m>1\}} + 2\lambda 1_{\{m=1\}}\big]&p(m-1,n) \qquad (1)\end{aligned}$$

where $\mu = \lambda + \gamma$, and $1_{\{\}}$ is the indicator function.

We define the joint probability generating function

$$P(x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p(m,n)x^m y^n, \quad |x| \le 1, |y| \le 1 \qquad (2)$$

and the generating functions for the boundary probabilities

$$\begin{aligned} P(x,0) &= \sum_{m=0}^{\infty} p(m,0)x^m, \quad |x| \le 1\\ P(0,y) &= \sum_{n=0}^{\infty} p(0,n)y^n \quad , \quad |y| \le 1 \qquad (3)\end{aligned}$$

From the balance equations it can be shown that $P(x,y)$ satisfies the following functional equation

$$k(x,y)P(x,y) = a(x,y)P(x,0) + b(x,y)P(0,y) \qquad (4)$$

where

$$\begin{aligned} k(x,y) &= rx(1-y) + \mu y(y-x) - \lambda x(1-x)y\\ a(x,y) &= rx(1-y)\\ b(x,y) &= \mu(y-x)y + \lambda x(1-x)y \end{aligned}$$

The function $k(x,y)$ is the so called *kernel* of the functional equation. Due to the regularity properties of $P(x,y)$, for each pair (x,y) on or within the unit circle for which the kernel $k(x,y) = 0$, the right hand side of (4) must vanish, i.e.

$$a(x,y)P(x,0) + b(x,y)P(0,y) = 0 \qquad (5)$$

Hence, additional relations between the unknown functions $P(x,0)$ and $P(0,y)$ can be obtained from examining the algebraic curve $k(x,y) = 0$ in the whole complex plane. In particular, the condition (5) can be formulated as a boundary value problem as first proposed in [12] and further developed in [13]–[15].

The kernel $k(x,y)$ is a polynomial of degree 2 in y. We thus have that for each value of x there are two possible values of y, say $y_1(x)$ and $y_2(x)$, such that $k(x,y_1(x)) = k(x,y_2(x)) = 0$. These functions are given by

$$y(x) = \frac{s(x) \pm \sqrt{s(x)^2 - 4\mu r x}}{2\mu} \qquad (6)$$

where $s(x) = x[r + \mu + \lambda(1-x)]$.

Furthermore, the algebraic function $y(x)$ has 4 real branch points $0 = x_1 < x_2 \le 1 < x_3 < x_4$, which are the zeros of the discriminant $D(x) = s(x)^2 - 4\mu r x$ of $k(x,y) = 0$. This follows from the fact that $D(0) = 0$, $D(x) < 0$ when x is very small, $D(1) \ge 0$, $D(x) < 0$ for $x = 1 + \frac{r+\mu}{\lambda}$, and $\lim_{x\to\infty} D(x) = \infty$.

For each $x \in [0, x_2]$, the two roots $y_1(x)$ and $y_2(x)$ are complex conjugates since $D(x)$ is zero for $x = 0$ and $x = x_2$ and negative for $x \in (0, x_2)$. Hence, the interval $[0, x_2]$ is mapped by $x \mapsto y(x)$ onto a closed contour L which is symmetric with respect to the real line and defined by $|y|^2 = \frac{xr}{\mu}$.

We will now describe how the curve L can be used to formulate a boundary value problem for the function $P(0,y)$. Since the function $P(x,0)$ is real for $x \in [0, x_2]$, we can rewrite the condition (5) as

$$\text{Im}\left[-\frac{b(\mu|y|^2/r, y)}{a(\mu|y|^2/r, y)}P(0,y)\right] = 0, \quad y \in L \qquad (7)$$

which is a homogenous Riemann-Hilbert boundary value problem with index $\chi = 0$ [16]. Define the following

$$-\frac{b(\mu|y|^2/r, y)}{a(\mu|y|^2/r, y)} \equiv \beta(y) + j\alpha(y), \;\; P(0,y) \;\equiv u(y) + jv(y)$$

and let $\hat{\alpha}(y) = \alpha(y)/\sqrt{\alpha(y)^2 + \beta(y)^2}$ and $\hat{\beta}(y) = \beta(y)/\sqrt{\alpha(y)^2 + \beta(y)^2}$. We can then rewrite the boundary condition of the problem as:

$$\hat{\alpha}(y)u(y) + \hat{\beta}(y)v(y) = 0, \quad y \in L \qquad (8)$$

The standard way of solving this boundary value problem is to transform the boundary condition to a condition on the unit circle C using conformal mapping. Define the conformal map $z = f(y) : L^+ \mapsto C^+$ and its inverse $y = f_0(z) : C^+ \mapsto L^+$ where $+$ indicates the region interior of the curve. Using these mappings we can reduce the Riemann-Hilbert problem on L to that on the unit circle C. Hence, the new problem is to determine a function $G(z) = u(z) + jv(z)$ such that

$$\hat{\alpha}(f_0(z))u(z) + \hat{\beta}(f_0(z))v(z) = 0, \quad z \in C \qquad (9)$$

which has the following solution [13]

$$G(z) = K \exp\left[\frac{1}{2\pi}\int_{t\in C} \arctan\left\{\frac{\hat{\beta}(f_0(t))}{\hat{\alpha}(f_0(t))}\right\}\frac{t+z}{t-z}\frac{dt}{t}\right] \qquad (10)$$

where K is a constant. The solution of the boundary value problem (7) is then given by [13]

$$P(0,y) = G(f(y)), \qquad y \in L^+ \qquad (11)$$

The conformal map $z = f(y) : L^+ \mapsto C^+$ can be found in closed-form [17] while the inverse mapping $y = f_0(z) : C^+ \mapsto L^+$ can be determined numerically using Theodorsen's procedure [13]. Hence, the function $P(0,y)$ has been determined; subsequently, we can substitute first in (5) to determine $P(x,0)$, then in (4) to finally obtain $P(x,y)$.

IV. Decomposition Approximation

Now we consider a more general and realistic model in which arrival rates may not be equal and transmission time is directly proportional to packet length which we assume constant. Since an exact analysis of the transmission queue under such assumptions is rather difficult, we will instead obtain an approximate solution using the decomposition technique in [18] which will be based on the unjustified assumption that the departure process from the coding stage is a renewal process. Another motivation for seeking a simple approximation is that the boundary value problem technique discussed previously may not be practical for optimizing performance in real time. The accuracy of the approximation will be verified via discrete event simulation with NS-2 [19].

A. Coding Queues

We represent the state of the two coding queues at time instant t by a single number $\hat{Q}_c(t) = Q_c^1(t) - Q_c^2(t) \in [-\infty, \infty]$. Let $p_c(m)$ be its stationary probability $p_c(m) = \lim_{t\to\infty} P[\hat{Q}_c(t) = m]$ which is given by

$$p_c(m) = \begin{cases} p_c(0)\left(\frac{\lambda_1}{\mu_1}\right)^m & , m > 0 \\ p_c(0)\left(\frac{\lambda_2}{\mu_2}\right)^{-m} & , m < 0 \end{cases} \tag{12}$$

where $\mu_1 = \gamma_1 + \lambda_2$ and $\mu_2 = \gamma_2 + \lambda_1$. The value of $p_c(0)$ can be obtained from the normalization condition $\sum_{m=-\infty}^{\infty} p_c(m) = 1$, yielding $p_c(0) = \left[1 + \lambda_1(\mu_1 - \lambda_1)^{-1} + \lambda_2(\mu_2 - \lambda_2)^{-1}\right]^{-1}$

Consider the queue length process at time instants when packets depart from the coding stage, and denote by $q_c(m)$ the stationary probability associated with the queue length at those instants which can be shown to have the following form

$$q_c(m) = \begin{cases} q_c(0)\rho_1\left(\frac{\lambda_1}{\mu_1}\right)^m & , m > 0 \\ q_c(0)\rho_2\left(\frac{\lambda_2}{\mu_2}\right)^{-m} & , m < 0 \end{cases} \tag{13}$$

where $q_c(0) = \left[1 + \rho_1\lambda_1(\mu_1 - \lambda_1)^{-1} + \rho_2\lambda_2(\mu_2 - \lambda_2)^{-1}\right]^{-1}$.

The distribution of the response time for the ith coding queue $T_c^i(x) = P[\mathcal{T}_c^i \le x]$, $i = 1, 2$ is given by

$$\begin{aligned} T_c^i(x) &= \sum_{n=\mp 1}^{\mp\infty} p_c(n) + \int_0^x \sum_{n=0}^{\pm\infty} p_c(n) \frac{\mu_i^{n+1} t^n e^{-\mu_i t}}{n!}\, dt \\ &= 1 - p_c(0)\frac{\mu_i}{\mu_i - \lambda_i}\, e^{-(\mu_i - \lambda_i)x}, \qquad x \ge 0 \end{aligned} \tag{14}$$

Note that the distribution of the response time is exponential with a jump of size $1 - p(0)\frac{\mu_i}{\mu_i - \lambda_i}$ at the origin. By straight forward calculations, the mean response time in the ith coding queue is

$$T_c^i = p_c(0)\frac{\mu_i}{(\mu_i - \lambda_i)^2} \tag{15}$$

We will approximate the departure process $\mathcal{D}_c$ from the coding stage using the stationary interval method which ignores the correlation between successive departures. The approximate distribution function $D_c(x) = P[\mathcal{D}_c < x]$ can then be expressed as

$$D_c(x) = \sum_{n=1}^{\infty} q_c(n)[1 - e^{-\mu_1 x}] + \sum_{n=-1}^{-\infty} q_c(n)[1 - e^{-\mu_2 x}]$$
$$+ q_c(0)\int_0^x [1 - e^{-\Lambda(x-y)}][\rho_1\mu_1 e^{-\mu_1 y} + \rho_2\mu_2 e^{-\mu_2 y}]dy$$

The mean output rate (throughput) from the coding stage is $S = \mathbb{E}[\mathcal{D}_c]^{-1}$ which can be shown to be

$$S = \Lambda - S_c = \Lambda - \frac{\lambda_1\lambda_2(\gamma_1 + \gamma_2)}{\lambda_1\gamma_1 + \lambda_2\gamma_2 + \gamma_1\gamma_2} \tag{16}$$

where S_c denotes the output rate of coded packets. We can see that $S < \Lambda$ for all values of $\gamma_i < \infty$, i.e., when network coding is applied.

B. Transmission Queue

We apply the diffusion approximation in [20] for a $G/D/1$ system in order to obtain the following expression for the mean response time of the transmission queue

$$T_{tr} \approx \frac{1}{r(1 - \hat{\varrho})} \tag{17}$$

where $\hat{\varrho} = \exp\left[\frac{-2(1-\varrho)}{K_d}\right]$, $\varrho = \frac{S}{r}$ and K_d is the square of the variation coefficient of $\mathcal{D}_c$.

We can summarize the *stability conditions* for the coding and transmission queues as follows:

$$\gamma_i > \lambda_i - \lambda_{3-i} \quad \text{and} \quad S < r \tag{18}$$

which for the balanced system reduce to

$$\begin{aligned} &\gamma > 0, && \text{if} \quad \Lambda < r \\ &0 < \gamma < \frac{\Lambda(r - \lambda)}{\Lambda - r}, && \text{if} \quad r \le \Lambda < 2r \end{aligned}$$

V. Optimizing the Time-out Parameters

We now use the analytical results derived previously in order to investigate the trade-offs associated with the choice of the time-out parameters. The performance measures of interest are delay, energy consumption and bandwidth utilization.

A. Response Time

Given a set of input rates $\{\lambda_1, \lambda_2\}$ and the transmission rate of the output link r we can determine the optimal time-out parameters that minimize the approximate expression for the mean response time of the system. To this end, we identify 3 traffic regimes within the stability region (18)

1) Light traffic conditions ($\Lambda/r \to 0$): coding cannot improve delay performance and therefore forwarding packets without coding is optimal.
2) Moderate traffic conditions ($\Lambda/r < 1$): the system is stable without using network coding. Nevertheless, coding can reduce both delay and transmission costs.
3) Heavy traffic conditions ($1 \le \Lambda/r < 2$): one needs to perform network coding in order to stabilize the transmission queue (i.e., to achieve $S < r$).

The mean response times of class 1 and 2 packets are not independent, therefore we construct the following aggregate objective function:

$$T = wT_c^1 + (1-w)T_c^2 + T_{tr} \tag{19}$$

where $w \in [0,1]$. We take $w = 0.5$ in order to give equal priority to both packet classes. The constrained optimization problem can thus be formulated as follows

$$\min_{\gamma_1,\gamma_2} T, \qquad \text{subject to stability conditions (18).}$$

B. Energy and Bandwidth

Energy consumption is an important performance measure in wireless environments especially in sensor networks where nodes are typically resource-constrained and rely on batteries or energy harvesting. Sensor nodes may also use dedicated hardware to implement each distinct functionality. Denote by ϵ_f the energy consumption per packet due to processing at the input queue, and let ϵ_c and ϵ_{tr} denote the energy consumed per a coding operation and the energy required to transmit a single packet, respectively. We can then compute the average *energy per unit time* consumption of the relay node when it only forwards packets

$$\mathcal{E}_f = \Lambda(\epsilon_f + \epsilon_{tr}) \tag{20}$$

and when it uses network coding

$$\mathcal{E}_c = \Lambda\epsilon_f + S_c\epsilon_c + (\Lambda - S_c)\epsilon_{tr} = \mathcal{E}_f - S_c(\epsilon_{tr} - \epsilon_c) \tag{21}$$

Hence, we have the following intuitive result

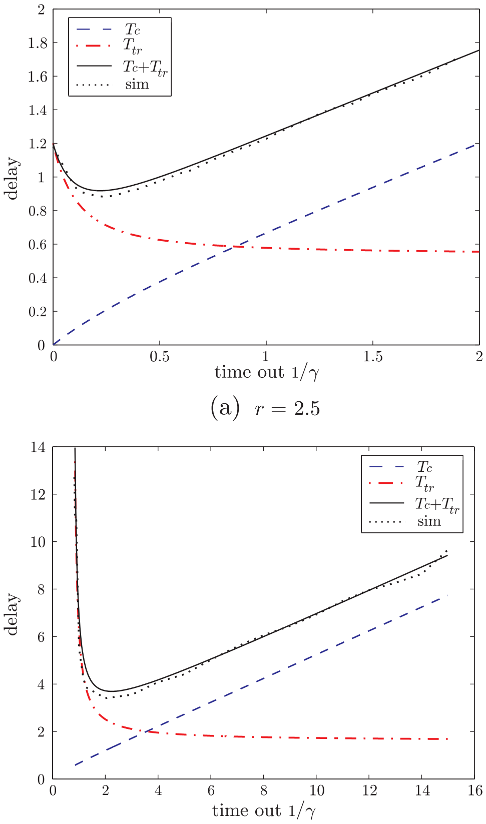
$$\mathcal{E}_c < \mathcal{E}_f, \quad \text{if} \quad \epsilon_c < \epsilon_{tr} \tag{22}$$

In order to measure the bandwidth savings offered by network coding, we define *coding gain* as the ratio of the input traffic rate to the output traffic rate, $\eta = \Lambda/S \in [1,2]$. The problem of maximizing the coding gain, however, has a trivial solution $\gamma_i^* = 0$, which renders the coding stage unstable. Hence, there is a compromise to be achieved between delay and coding gain which will be illustrated in the numerical results.

VI. Numerical Examples

In this section, we present numerical results which illustrate the different trade-offs under various traffic conditions. We use the proposed approximation for constant transmission times, and we validate the accuracy of the model via simulations.

Fig. 2(a) depicts the mean coding, transmission and total delays in the system vs the mean time-out period for a balanced system under moderate traffic condition, $\Lambda = 2$ and $r = 2.5$. In this case, the range of time-out parameters which stabilizes the system is $[0,\infty)$ where the lower bound corresponds to the non-coding case. The figure indicates that the approximation yields very accurate results as compared to simulations. We can also observe that the coding delay is monotonically increasing while the transmission delay is exponentially decaying with the time-out period. Hence, the total delay in the system decreases up to a point (the optimal) after which it increases continuously. More specifically, network coding reduces the average response time by up to 24% as compared to plain forwarding while at the same time offering a coding gain of 1.3. Furthermore, a coding gain of 1.5 can be achieved while maintaining a similar response time to the non-coding system.



(a) $r = 2.5$

(b) $r = 1.4$

Fig. 2. Average coding delay T_c, transmission delay T_{tr} and the total delay vs the mean time-out interval $1/\gamma$ for balanced traffic with $\Lambda = 2$.

Fig. 2(b) depicts similar numerical results under heavy traffic conditions where forwarding packets without coding renders the transmission queue unstable. The parameters used in the figure are $\Lambda = 2$ and $r = 1.4$, which imply that the average time-out periods can vary in the range $(0.75,\infty)$ in order to stabilize the system. In this case, network coding yields a minimum response time of about 3.7 s while offering a coding gain of 1.7. Higher coding gains can also be achieved at the expense of higher delays.

Fig. 3 depicts analytical results for the trade-off between delay and coding gain as a result of varying the time-out interval for a balanced system under different traffic conditions. From the figure, we can conclude that the time-out period should not be less than the optimal value since the trade-off curve is almost symmetric around this point. The figure also suggests that the range of time-out intervals which can stabilize the system decreases as the traffic rates increase.

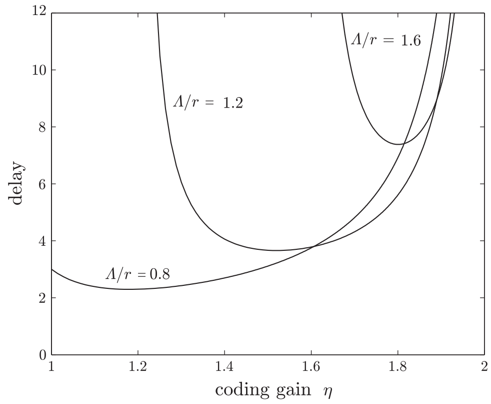


Fig. 3. The trade-off between delay and coding gain as a result of varying the time-out period for a balanced system under different traffic conditions.

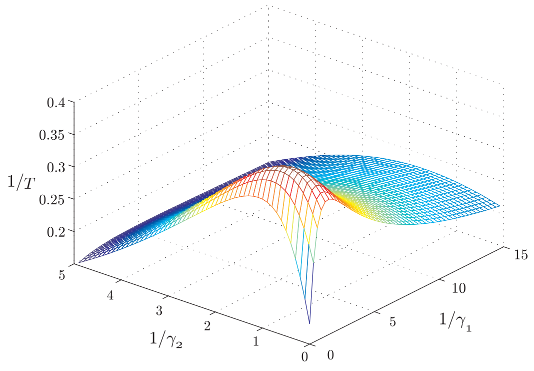


Fig. 4. The inverse of the aggregate delay function T^{-1} vs the mean time-out periods for $\lambda_1 = 0.4$, $\lambda_2 = 0.5$ and $r = 1$.

In Fig. 4, we plot the inverse of the aggregate delay function (19) when the arrival rates of the two classes are moderately different. The figure shows that there is an optimal set of time-out parameters which maximizes the inverse of the delay cost function, yielding a delay improvement of about 54% and a coding gain of 1.26 in comparison to plain forwarding. Finally, we have observed that when the node is moderate to heavily loaded and the arrival rates of the two classes are highly unbalanced (i.e. $\lambda_j/\lambda_i << 1$), the optimization problem depends mainly on the time-out parameter for the faster flow γ_i while the optimal value for the slower flow is $\gamma_j^* = 0$. However, we have not been able to verify this observation mathematically.

VII. Conclusions

In this paper, we have proposed and analyzed a queueing model that captures the different functionalities of a network coding router including packet classification, route processing, coding and transmission. We assumed that the router employs a time-out mechanism in order to accumulate packets for coding, and we evaluated the trade-offs associated with varying the length of the waiting time. We have shown that an optimal value of the time-out interval exists that minimizes the average response time of the system. This work has focused on the two-flow case; further work is required in order to investigate and model the trade-offs for arbitrary number of flows. Future work will also extend the analysis to a multi-hop setting where performance metrics such as the end-to-end delay including packet assembly and decoding at the output are also considered.

Acknowledgment

This research was supported by the EU FP7 DIESIS project.

References

[1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] X. He and A. Yener, "On the energy-delay trade-off of a two-way relay network," in *Proc. 42nd Annual Conference on Information Sciences and Systems (CISS '08)*, Princeton, NJ, USA, Mar. 2008, pp. 865–870.

[3] W. Chen, K. B. Letaief, and Z. Cao, "Opportunistic network coding for wireless networks," in *Proc. IEEE International Conference on Communications (ICC '07)*, Glasgow, Scotland, 2007, pp. 4634–4639.

[4] D. Umehara, T. Hirano, S. Denno, M. Morikura, and T. Sugiyama, "Wireless network coding in slotted aloha with two-hop unbalanced traffic," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 647–661, Jun. 2009.

[5] Y. Sagduyu and A. Ephremides, "Cross-layer optimization of mac and network coding in wireless queueing tandem networks," *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 554–571, Feb. 2008.

[6] ——, "On joint mac and network coding in wireless ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3697–3713, Oct. 2007.

[7] P. Parag and J.-F. Chamberland, "Queueing analysis of a butterfly network," in *Proc. IEEE International Symposium on Information Theory (ISIT '08)*, Ontario, Canada, Jul. 2008, pp. 672–676.

[8] A. Mahmino, J. Lacan, and C. Fraboul, "Guaranteed packet delays with network coding," in *Proc. 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '08)*, San Francisco, CA, USA, Jun. 2008, pp. 1–6.

[9] O. H. Abdelrahman and E. Gelenbe, "Queueing performance under network coding," in *Proc. IEEE Information Theory Workshop (ITW '09)*, Volos, Greece, Jun. 2009, pp. 135–139.

[10] ——, "Approximate analysis of a round robin scheduling scheme for network coding," in *Proc. 6th European Performance Engineering Workshop (EPEW '09)*, ser. LNCS, vol. 5652, London, UK, Jul. 2009, pp. 212–217.

[11] D. E. Lucani, M. Medard, and M. Stojanovic, "Random linear network coding for time-division duplexing: Queueing analysis," in *Proc. IEEE International Symposium on Information Theory (ISIT '09)*, Seoul, Korea, Jul. 2009, pp. 1423–1427.

[12] G. Fayolle and R. Iasnogorodski, "Two coupled processors: The reduction to a Riemann-Hilbert problem," *Probability Theory and Related Fields*, vol. 47, no. 3, pp. 325–351, Jan. 1979.

[13] J. W. Cohen and O. J. Boxma, *Boundary value problems in queueing system analysis*. Amsterdam ; New York: North-Holland Pub. Co.; Elsevier Science Pub. Co., 1983.

[14] J. S. H. van Leeuwaarden and J. A. C. Resing, "A tandem queue with coupled processors: Computational issues," *Queueing Systems*, vol. 51, no. 1, pp. 29–52, Jan. 1979.

[15] A. M. Haghighi and D. P. Mishev, "Analysis of a two-node task-splitting feedback tandem queue with infinite buffers by functional equation," *International Journal of Mathematics in Operational Research*, vol. 1, pp. 246–277, Jan. 2009.

[16] F. D. Gakhov, *Boundary value problems*. Pergamon Press, 1966.

[17] J. P. C. Blanc, "The relaxation time of two queueing systems in series," *Stochastic Models*, vol. 1, no. 1, pp. 1–16, 1985.

[18] E. Gelenbe and G. Pujolle, *Introduction to queueing networks*. New York, NY, USA: John Wiley & Sons, Inc., 1987 and 2000.

[19] The Network Simulator (ns-2). http://www.isi.edu/nsnam/ns/.

[20] E. Gelenbe, "On approximate computer system models," *Journal of the ACM*, vol. 22, no. 2, pp. 261–269, Apr. 1975.