

Network With Costs: Timing and Flow Decomposition

Shreeshankar Bodas, Jared Grubb, Sriram Sridharan, Tracey Ho, Sriram Vishwanath
{bodas, grubb, sridhara, sriram}@ece.utexas.edu, tho@caltech.edu

Abstract—This paper analyzes a capacitated network with costs from an information theoretic point of view. Specifically, it accomplishes the following for a series-parallel single source-destination network:

1. Determines a flow decomposition for a network with costs starting from an information theoretic point of view.
2. Devises an algorithm that perform this flow decomposition for large packet sizes.
3. Establishes that timing plays a negligible role in capacity if packet size is sufficiently large.

I. INTRODUCTION

Networks with costs are abstractions that succinctly capture the underlying structure commercial wire-line networks. In this setting, the network is represented by a graph with non-interacting links, meaning that there is no interference and broadcast constraints on the links, with a cost associated with each link. Such a network is often called a *capacitated* network with costs, where the cost models the price (in dollars per use) associated with each link. Such a cost effectively models the setting where the network-manager bills each source-destination pair in proportion to the amount of time that pair utilizes the network. Inherent in such networks is a trade-off between cost and throughput, and understanding this trade-off is the main thrust of this paper.

In [16], Sergio Verdú derives the result in a similar set up for single link and for a MAC channel. In a recent result by Ralf Koetter [17], the author shows that, for any capacitated network (without costs), the capacity region (throughput maximization) problem reduces to a flow decomposition problem. Such a result is particularly powerful as it reduces an otherwise complex information theoretic problem into a well understood problem in networking [15]. In essence, Koetter shows that each link in the network can be replaced by an equivalent deterministic link with the same capacity while suffering no loss in performance.

In this paper, we consider a capacity network with costs. The channel is assumed to be a memoryless channel given by the conditional distribution $p(y|x)$, and the cost is assumed to be a *per use* cost, meaning that there is a penalty associated with each symbol being sent over a link in the network. Such a network differs from the cost-free network in the following ways:

1. Not using the link (the null symbol) has no cost, and thus *timing* can potentially be used to convey information on each link.

2. There is a cost to throughput tradeoff, meaning that the problem is no longer a pure throughput maximization problem, but has a trade-off between total cost and rate.

Our goal is to show the following for a single-source single-destination series-parallel network with costs:

1. The problem has a flow decomposition similar to that of cost-free networks when studied from an information theoretic perspective, regardless of the point of operation on the cost-throughput tradeoff curve. The network is thus insensitive to the actual transition probability $p(y|x)$ of the channel, and each link can be replaced by an equivalent deterministic link.
2. A step-by-step algorithm with $O(n)$ complexity (where n is the number of edges) that performs the flow decomposition for the network, determining the cost-throughput tradeoff within $O(n/m)$, where m is the packet size.
3. To establish that timing contributes at most $O(n/m)$ to the total throughput of the network.

A series-parallel network is a well defined concept from circuit theory. A series-parallel network is one where each sub-network of this network can be reduced to two component networks that are either connected in series or in parallel. Such networks form an important subclass of the set of all possible networks, and offers greater analytical tractability and intuition than an equivalent study of all possible capacitated networks.

Existing literature on networks with costs adopts a multitude of strategies to understand the performance of such networks [1][2][3][4] with a bulk of recent work on the topic using network coding [5][6][7][8][9][10][11]. However, in a majority of these results, it is assumed that no information is conveyed in a time slot if the transmitter chooses not to send a packet/symbol in that slot. A key issue studied in this paper is that of timing, i.e., the information conveyed when no packet is sent on the channel. As discussed in [12][13][14], timing comes in when the channel is allowed to have an idle state. For example, consider the following scenario in which a transmitter sends three packets over four time slots: the four time-slots can be utilized as (packet, silence, packet, packet) or (packet, packet, silence, packet), among other possibilities. Each of these utilizations incurs the same cost, because the channel is used only three times, but by clever sequencing of silence as in this case, it is possible to transmit more information than conveyed otherwise (that is, with only three packets). Timing information is typically difficult to analyze.

Our approach is to upper bound the contribution of timing to network throughput and establish that this impact is negligibly small as packet sizes increase to infinity.

Organization

The rest of the paper is organized as follows. Section II gives the system model. Section III considers the case when the system consists of a single channel from the source to the destination. In sections IV and V, we consider series and parallel networks of channels, respectively. Section VI presents an algorithm for finding the optimal transmission strategy for any given network that can be broken down into series and parallel links. We conclude with summary in section VII.

II. SYSTEM MODEL

A channel is a directed connection between a transmitter and a receiver and is characterized by a three-tuple (C, S, m) . The transmitter sends a packet over the channel, which takes m time-slots and can convey C bits per time-slot (for a total of mC bits per packet). The input and output alphabets are $\mathcal{X} = \mathcal{Y} = GF(2^m) \cup \{\Phi\}$, respectively. If the packet is Φ , then no cost is incurred in transmission. Otherwise, the channel incurs a cost S per time-slot (for a total cost of mS for the packet). The packet Φ corresponds to an idle state for the channel and will be referred to as the Null packet. The Null packet is special in that the receiver receives a Null packet if and only if a Null packet was sent (it can accurately detect whether the transmitter was idle or not). Furthermore, whenever a packet from $GF(2^m)$ is transmitted, then a (possibly different) symbol from $GF(2^m)$ is received.

A network is a directed acyclic graph $G(V, E)$, where V is the set of nodes in the network and E is the set of channels between nodes. We shall restrict our model to a special network topology: those that have a series and parallel recursive structure. These are networks that can be recursively decomposed into subgraphs such that every subgraph is either a pure-series or pure-parallel of subsequent subgraphs (until each subgraph at its most basic level is a single edge). An example of a network that does *not* fit our criterion is shown in Fig. 1.

We assume that the channels are independent of each other and memoryless. Further, the intermediate nodes can simultaneously transmit and receive.

We look at the problem of maximizing the rate of transmission from the source to the destination given a total average cost constraint on the network.

III. SOLUTION FOR A POINT-TO-POINT CHANNEL

In this section, we consider a point to point channel connecting the source and the destination. We solve for the maximum possible rate of transmission between the source and the destination subject to an average cost constraint. Let $W = \{1, 2, \dots, 2^{nmR}\}$ denote the set of message symbols. The source encodes the message into the codeword $X^n = [X_1, X_2, \dots, X_n]$ where $X_i \in \mathcal{X}$. The received message is denoted by Y^n , where $Y_i \in \mathcal{Y}$ and the channel is denoted by

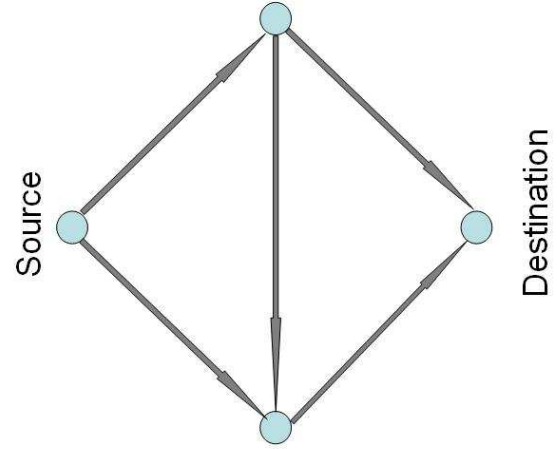


Fig. 1. A network that can NOT be represented as a combination of series and parallel links

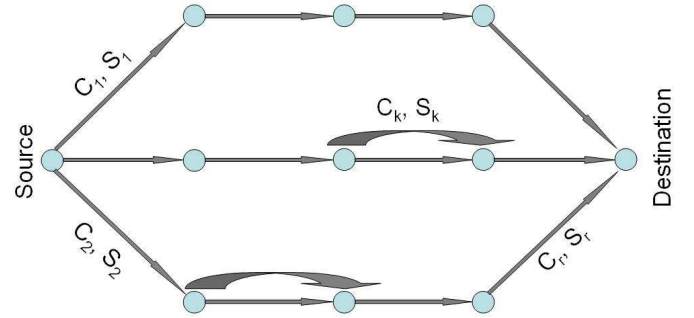


Fig. 2. A Typical Network

$(X^n, \prod_{i=1}^n p(Y_i|X_i), Y^n)$. Let (C, S, m) be the triplet characterizing this channel, S_0 the average cost per time slot constraint on the network, γ_i the probability that a packet other than Φ is transmitted in i^{th} packet-slot. By Fano's inequality, we can write the following upper bound on R :

$$nmR \leq I(X^n; Y^n) + n\epsilon \quad (1)$$

The capacity of the channel is $C = (1/m) \max_{p(X)} I(X; Y)$. Hence the optimization problem can be formulated as:

$$\text{Maximize } \lim_{n \rightarrow \infty} \frac{1}{mn} I(X^n; Y^n) \quad (2)$$

$$\text{subject to } \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \gamma_i \right) S \leq S_0 \quad (3)$$

Lemma 3.1: Let \tilde{C} denote the capacity of the channel when the Null symbol is not a part of the input alphabet. Then,

$$\tilde{C} \leq C \leq \tilde{C} + \frac{1}{m} \quad (4)$$

Proof: The first inequality, $\tilde{C} \leq C$, is immediate since $GF(2^m) \subset GF(2^m) \cup \{\Phi\}$. For the second inequality, let

$\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}} = GF(2^m)$ denote the input and output alphabets without the Null symbol.

$$\begin{aligned} I(X_i; Y_i) &= H(Y_i) - H(Y_i|X_i) \\ &= \sum_{\alpha \in GF(2^m) \cup \{\Phi\}} P(Y_i = \alpha) \log \frac{1}{P(Y_i = \alpha)} \\ &\quad - \sum_{\alpha \in GF(2^m) \cup \{\Phi\}} H(Y_i|X_i = \alpha) P(X_i = \alpha) \end{aligned}$$

Hence, replacing X_i and Y_i by \tilde{X}_i and \tilde{Y}_i respectively and since $H(Y_i|X_i = \Phi) = 0$, we get

$$\begin{aligned} I(X_i; Y_i) &= \sum_{\alpha \in GF(2^m)} \gamma_i P(\tilde{Y}_i = \alpha) \log \frac{1}{\gamma_i P(\tilde{Y}_i = \alpha)} \\ &\quad + P(Y_i = \Phi) \log \frac{1}{P(Y_i = \Phi)} \\ &\quad + \sum_{\alpha \in GF(2^m)} H(\tilde{Y}_i|\tilde{X}_i = \alpha) \cdot \gamma_i P(\tilde{X}_i = \alpha) \end{aligned}$$

Since $P(Y_i = \Phi) = 1 - \gamma_i$, $\gamma_i \leq 1$ and $H(\gamma_i) \leq 1$, we have

$$\begin{aligned} I(X_i; Y_i) &= H(\gamma_i) + \gamma_i I(\tilde{X}_i; \tilde{Y}_i) \\ &\leq 1 + \gamma_i I(\tilde{X}_i; \tilde{Y}_i) \\ &\leq 1 + I(\tilde{X}_i; \tilde{Y}_i) \end{aligned} \quad (5)$$

Transmission of each input symbol requires m channel uses. Therefore,

$$\begin{aligned} C &= \frac{1}{m} \max_{p(X)} I(X_i; Y_i) \\ &= \frac{1}{m} I(X_i; Y_i) \Big|_{p^*(X)} \\ &\leq \frac{1}{m} (1 + I(\tilde{X}_i; \tilde{Y}_i)) \Big|_{p^*(X)} \\ &= \frac{1}{m} + \frac{1}{m} I(\tilde{X}_i; \tilde{Y}_i) \Big|_{p^*(X)} \\ &\leq \frac{1}{m} + \max_{p(\tilde{X}_i)} \frac{1}{m} I(\tilde{X}_i; \tilde{Y}_i) \\ &= \frac{1}{m} + \tilde{C} \end{aligned}$$

The last 3 expressions are to be understood as follows: The channel capacity C equals the mutual information between X_i and Y_i for a certain input distribution $p^*(X)$. This distribution induces a distribution for \tilde{X}_i and correspondingly, there's a certain mutual information between the two random variables under question. This (induced) distribution may not be the "capacity-achieving" distribution, hence the second-last inequality follows. ■

Theorem 3.2: The capacity of the point to point channel under cost constraint, C_{p-p} is bounded by

$$(C - 1/m) \cdot \min(1, S_0/S) \leq C_{p-p} \leq \min(C, CS_0/S + 1/m).$$

Proof: The upper bound on C_{p-p} is straightforward, since $C_{p-p} \leq C$ and, from Fano's inequality we have

$$\begin{aligned} nmR &\leq I(X^n; Y^n) + n\epsilon \\ &\leq \sum_{i=1}^n I(X_i; Y_i) + n\epsilon \\ &\leq \sum_{i=1}^n (1 + \gamma_i I(\tilde{X}_i; \tilde{Y}_i)) + n\epsilon \\ \Rightarrow R &\leq \tilde{C} \cdot (S_0/S) + (1 + \epsilon)/m \end{aligned}$$

The second inequality follows because conditioning reduces entropy and by the memoryless nature of channels, the third one from eq. (5), and the last step follows because $\frac{1}{n} \sum_{i=1}^n \gamma_i \leq S_0/S$ by cost constraint, and $I(\tilde{X}_i; \tilde{Y}_i) \leq \tilde{C}$. Finally, since the last equation must hold for all $\epsilon > 0$, we get $C_{p-p} \leq \tilde{C} \cdot (S_0/S) + 1/m \leq CS_0/S + 1/m$.

For the lower bound, it is enough to prove that for any given $\epsilon > 0$, there exists a transmission strategy that achieves a rate greater than or equal to $(\tilde{C} - \epsilon) \min(1, S_0/S)$, implying $C_{p-p} \geq \tilde{C} \min(1, S_0/S)$. Since $\tilde{C} \geq C - 1/m$, the lower bound follows.

Proposed strategy: Define $\gamma_i = \gamma = \min(1, S_0/S)$ where, as before, γ_i is the probability that a non-Null packet is transmitted in i^{th} time slot. This scheme evidently obeys the cost constraint and so, is feasible.

The capacity of the channel when the Null symbol is not a part of the input alphabet is \tilde{C} , that is, every packet can communicate a maximum of $m\tilde{C}$ bits of information. Hence, for any given $\epsilon > 0$, there exists n_0 such that for all $n > n_0$, there exists a $(2^{nmR}, nm)$ codebook where $R = \tilde{C} - \epsilon$ is the achieved rate. Thus, the rate achieved under the proposed scheme is $\gamma R = \min(1, S_0/S) \cdot (\tilde{C} - \epsilon)$. It follows that $C_{p-p} \geq \min(1, S_0/S) \cdot (\tilde{C} - \epsilon)$. Therefore, $C_{p-p} \geq \tilde{C} \min(1, S_0/S)$ because $\epsilon > 0$ is arbitrary to begin with. ■

For large packet sizes (large m), a typical curve of achievable rate v/s allowed cost per time-slot will look like Fig. 3.

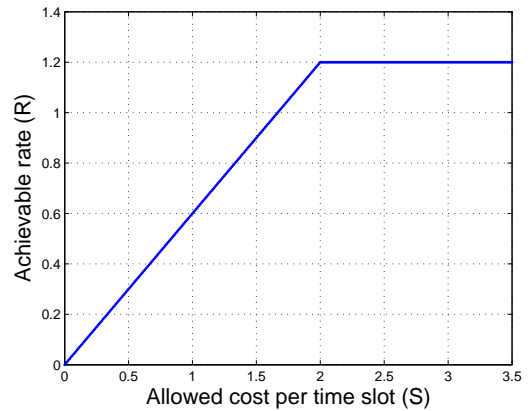


Fig. 3. Point-to-point link

IV. EXTENDING THE RESULT TO A SERIES OF CHANNELS

We consider a series of k channels connecting the source and the destination. The channels are memoryless, independent of each other and are characterized by the parameters (C_i, S_i, m_i) , $1 \leq i \leq k$, where (C_j, S_j, m_j) are the parameters of the j^{th} channel from the source node. We denote the input and output of the i^{th} channel by X_i and Y_i respectively. Thus $X_i, Y_i \in GF(2^{m_i}) \cup \{\Phi\}$.

Proposition 4.1: In the absence of cost constraint, the capacity of the series link is given by

$$C_{\text{series}} = \min\{C_1, C_2, \dots, C_k\}.$$

The proof follows from the data processing inequality applied to the Markov chain $X_1 \rightarrow X_i \rightarrow Y_i \rightarrow Y_k$ and has been omitted here.

Theorem 4.2: The capacity of the series link under the cost constraint S_0 is bounded by

$$C_{\text{series}} \geq \min\left(C_r - \frac{1}{m_r}, \frac{S_0}{\frac{S_1}{C_1 - \frac{1}{m_1}} + \frac{S_2}{C_2 - \frac{1}{m_2}} + \dots + \frac{S_k}{C_k - \frac{1}{m_k}}}\right),$$

$$C_{\text{series}} \leq \min\left(C_r, \frac{S_0}{\frac{S_1}{C_1 + \frac{1}{m_1 \xi_1}} + \frac{S_2}{C_2 + \frac{1}{m_2 \xi_2}} + \dots + \frac{S_k}{C_k + \frac{1}{m_k \xi_k}}}\right)$$

where $r = \operatorname{argmin}_{1 \leq i \leq k} C_i$, and ξ_i is the probability (or the fraction of time, for large time blocks) that the i^{th} channel is used.

Proof:

Consider a time equivalent to N time-slots, N is large. We solve the following optimization problem:

$$\text{Maximize } \lim_{N \rightarrow \infty} \frac{1}{N} I(X_1^{n_1}; Y_k^{n_k})$$

subject to:

- 1) $0 \leq \gamma_{i,j} \leq 1$ for all $1 \leq i \leq k$, $1 \leq j \leq n_i$
- 2) $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^k (\sum_{j=1}^{n_i} \gamma_{i,j}) S_i \leq S_0$ (The average cost constraint)

Here, $n_i = N/m_i$ is the possible number of symbols (packets) that can be transmitted on i^{th} channel in time equivalent to N time-slots, $\gamma_{i,j}$ is the probability that a valid (non-Null) packet is transmitted on i^{th} channel in j^{th} symbol-slot. The problem is to choose the optimal values of $\gamma_{i,j}$ for $1 \leq i \leq k$, $1 \leq j \leq n_i$. As before, \tilde{C}_i denotes the capacity of the channel when the Null symbol is not a part of the input alphabet.

Before continuing with the proof, we prove the following claim that will be useful in the proof.

Claim: $C_{\text{series}} \leq (\frac{1}{n_i} \sum_{j=1}^{n_i} \gamma_{i,j}) \tilde{C}_i + \frac{1}{m_i}$ for all $i \in \{1, 2, \dots, k\}$.

Proof: Let $\xi_i = (\frac{1}{n_i} \sum_{j=1}^{n_i} \gamma_{i,j})$. From Fano's inequality, we can write, for an end-to-end rate R over the series assembly,

$$\begin{aligned} NR &\leq I(X_1^{n_1}; Y_k^{n_k}) + N\epsilon \\ &\leq \min_i (I(X_i^{n_i}; Y_i^{n_i})) + n\epsilon \\ &\leq \min_i \sum_{j=1}^{n_i} (1 + \gamma_{i,j} I(\tilde{X}_{i,j}; \tilde{Y}_{i,j})) + N\epsilon \\ \Rightarrow R &\leq \min_i \left(\frac{1}{m_i} + \epsilon + \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \gamma_{i,j} \right) \tilde{C}_i \right) \end{aligned}$$

The second inequality, again, follows because conditioning reduces entropy and the memoryless nature of the channels, the third inequality follows from eq. (5). Since the last inequality must hold for all $\epsilon > 0$, we get the desired result.

[Returning to the proof of the theorem] The cost of transmission incurred over N time-slots and all the channels put together is $S = N(\xi_1 S_1 + \xi_2 S_2 + \dots + \xi_k S_k)$ and the amount of data transferred (end-to-end) is $N \times C_{\text{series}}$. Thus, the cost constraint implies

$$\begin{aligned} S_0 &\geq \xi_1 S_1 + \xi_2 S_2 + \dots + \xi_k S_k \\ \Leftrightarrow 1 &\leq S_0 / (\xi_1 S_1 + \xi_2 S_2 + \dots + \xi_k S_k) \\ \Leftrightarrow C_{\text{series}} &\leq S_0 / \left(\frac{\xi_1 S_1}{C_{\text{series}}} + \frac{\xi_2 S_2}{C_{\text{series}}} + \dots + \frac{\xi_k S_k}{C_{\text{series}}} \right) \end{aligned}$$

From the previous claim, we have $C_{\text{series}} \leq \xi_i \tilde{C}_i + 1/m_i$

$$\Rightarrow C_{\text{series}} \leq S_0 / \left(\frac{S_1}{C_1 + \frac{1}{m_1 \xi_1}} + \frac{S_2}{C_2 + \frac{1}{m_2 \xi_2}} + \dots + \frac{S_k}{C_k + \frac{1}{m_k \xi_k}} \right)$$

Since we already have $C_{\text{series}} \leq C_r$, the upper bound on C_{series} has been proved and we need only consider the lower bound.

Case 1: $\tilde{C}_r (\sum_{i=1}^k S_i / \tilde{C}_i) \leq S_0$.

In this case, $\gamma_{i,j} = \tilde{C}_r / \tilde{C}_i$ is achievable. The rate achievable under this scheme is \tilde{C}_r . Since $\tilde{C}_i \geq C_i - 1/m_i \forall i$, the lower bound follows.

Case 2: $\tilde{C}_r (\sum_{i=1}^k S_i / \tilde{C}_i) > S_0$.

Consider the following scheme for operating the channels:

$$\gamma_{i,j} = \xi_i = \frac{S_0 / \tilde{C}_i}{(S_1 / \tilde{C}_1 + S_2 / \tilde{C}_2 + \dots + S_k / \tilde{C}_k)}.$$

This scheme obeys the cost constraint, since the overall cost incurred over N time-slots equals $N(\xi_i S_i + \xi_2 S_2 + \dots + \xi_k S_k) \leq N S_0$. Substituting the value of ξ_i , we verify that the constraint is met with equality.

The end-to-end rate achieved by this scheme equals

$$\min_{1 \leq i \leq k} (\xi_i C_i) = \frac{S_0}{(S_1 / \tilde{C}_1 + S_2 / \tilde{C}_2 + \dots + S_k / \tilde{C}_k)}$$

This rate (and therefore, C_{series}) is no smaller than $S_0 / (\frac{S_1}{C_1 - \frac{1}{m_1}} + \frac{S_2}{C_2 - \frac{1}{m_2}} + \dots + \frac{S_k}{C_k - \frac{1}{m_k}})$, since $\tilde{C}_i \geq C_i - 1/m_i$. Hence, the lower bound follows. ■

We note that for any given cost constraint, for large packet sizes (large m_i), the upper and lower bounds match. For large packet sizes, a typical plot of the achievable rate over the series of channels v/s allowed cost per time-slot is shown in Fig. 4.

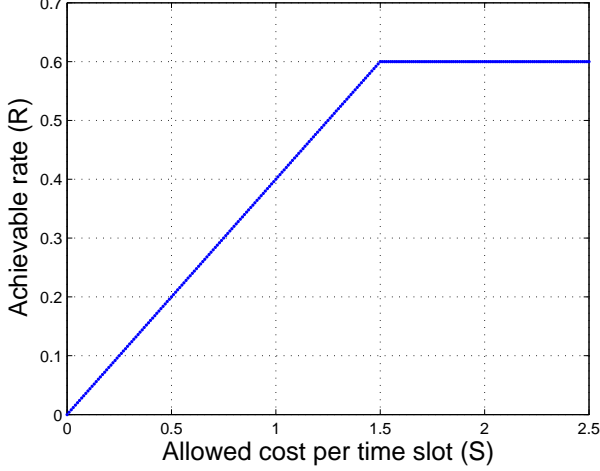


Fig. 4. Series Link

V. PARALLEL CHANNELS

Consider the case of a source and a sink connected by a set of k point-to-point channels in parallel; that is, the head and tail of each channel is at the source and sink, respectively. The capacity-cost-packet triplet for channel i is denoted by (C_i, S_i, m_i) .

For each channel i and “packet-slot” j , the source may either send a random variable $X_{i,j} \in \{1, \dots, 2^{m_i}\}$ (incurring cost $m_i S_i$) or not use the channel (sending Φ_i and incurring zero cost). So, by time N , a channel i has transmitted $\frac{N}{m_i}$ packets. Let $\mathbf{X}^N := \bigoplus_{i=1}^k (X_{(i,1)}, \dots, X_{(i, \frac{N}{m_i})})$ be the sequence of all packets sent on the parallel network. The destination, in turn, receives the vector $\mathbf{Y}^n := \bigoplus_{i=1}^k (Y_{(i,1)}, \dots, Y_{(i, \frac{N}{m_i})})$. Therefore, the source can encode a message $\mathbf{W} \in (1, \dots, 2^{\sum_{i=1}^k \frac{N}{m_i} R})$ and we seek to maximize R .

As before, let $\gamma_{i,j}$ represent the probability that channel i transmits a non-Null symbol in time-slot j . Our objective is to solve the following optimization problem:

$$C_{par} := \lim_{N \rightarrow \infty} \max_{\{\gamma_{i,j}\}} \frac{1}{N} I(\mathbf{X}^N; \mathbf{Y}^N)$$

subject to the cost constraint,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j} m_i S_i \leq S_0$$

We note that

$$\begin{aligned} I(\mathbf{X}^N; \mathbf{Y}^N) &\leq \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} I(X_{i,j}; Y_{i,j}) \\ &\leq \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \left(1 + \gamma_{i,j} I(\tilde{X}_{i,j}; \tilde{Y}_{i,j})\right) \\ &\leq \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \left(1 + \gamma_{i,j} m_i \tilde{C}_i\right) \\ &\leq \sum_{i=1}^k \frac{N}{m_i} + \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j} m_i \tilde{C}_i \\ \frac{1}{N} I(\mathbf{X}^N; \mathbf{Y}^N) &\leq \sum_{i=1}^k \frac{1}{m_i} + \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j} m_i \tilde{C}_i \end{aligned}$$

The first step follows from the memoryless nature of the channels. As defined earlier, $\tilde{X}_{i,j}$ and $\tilde{Y}_{i,j}$ are the non-Null symbols transmitted; the 1 in the second inequality arises as before from $H(\gamma_{i,j}) \leq 1$.

Given some complete policy $\bigoplus_{i=1}^k (\gamma_{i,1}, \dots, \gamma_{i, \frac{N}{m_i}})$, let $(\gamma_i)_{i=1, \dots, k}$ be a time-invariant policy, where each component is the computed average, namely $\gamma_i := \frac{m_i}{N} \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j}$. Our optimization problem then is:

$$\begin{aligned} C_{par} &\leq \max_{(\gamma_{i,j})} \lim_{N \rightarrow \infty} \left(\sum_{i=1}^k \frac{1}{m_i} + \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j} m_i \tilde{C}_i \right) \\ &= \sum_{i=1}^k \frac{1}{m_i} + \max_{(\gamma_{i,j})} \lim_{N \rightarrow \infty} \sum_{i=1}^k \tilde{C}_i \left(\frac{m_i}{N} \sum_{j=1}^{\frac{N}{m_i}} \gamma_{i,j} \right) \\ &= \sum_{i=1}^k \frac{1}{m_i} + \max_{(\gamma_i)} \lim_{N \rightarrow \infty} \sum_{i=1}^k \gamma_i \tilde{C}_i \\ &= \sum_{i=1}^k \frac{1}{m_i} + \max_{(\gamma_i)} \sum_{i=1}^k \gamma_i \tilde{C}_i \end{aligned} \quad (6)$$

subject to the cost constraint,

$$\sum_{i=1}^k \gamma_i S_i \leq S_0$$

Note that the time-invariant policy (γ_i) and the complete policy $(\gamma_{i,j})$ provide the same upper bound on capacity. Therefore, we shall only consider time-invariant policies.

We shall assume that the channels are ordered $\frac{S_i}{\tilde{C}_i} \leq \frac{S_{i+1}}{\tilde{C}_{i+1}}$, an ordering from cheapest channel (per time-slot) to most expensive.

Theorem 5.1: Let R^* be the rate given by

$$R^* = \sum_{j=1}^k \gamma_j^* C_j$$

where k^* is determined so that $S_0 \in \left[\sum_{i=1}^{k^*-1} S_i, \sum_{i=1}^{k^*} S_i \right)$, and the terms $(\gamma_1^*, \dots, \gamma_k^*) := (1, \dots, 1, \gamma_{k^*}^*, 0, \dots, 0)$ and $\gamma_{k^*}^* := \frac{S_0 - \sum_{i=1}^{k^*-1} S_i}{S_{k^*}}$. Then the capacity of the parallel links under the cost constraint S_0 is bounded by

$$R^* - \sum_{i=1}^{k^*} \frac{1}{m_i} \leq C_{par} \leq R^* + \sum_{i=1}^k \frac{1}{m_i}.$$

(Note: If $S_0 \geq \sum_{i=1}^k S_i$, then we shall use the policy $\gamma := (\gamma_1, \dots, \gamma_k) = (1, 1, \dots, 1)$ since cost is not prohibitive. If $S_0 > 0$, then $\gamma = (\epsilon, 0, \dots, 0)$ with $\epsilon = \min\{1, \frac{S_0}{S_1}\}$ is better than the all-zero policy. We shall dismiss these degenerate cases and assume that $S_0 \in (0, \sum_{i=1}^k S_i)$, so that the entries are not all zero or not all one.)

Definition: Given a policy $\gamma = (\gamma_1, \dots, \gamma_k)$, let $h_\gamma := \min\{i : \gamma_i < 1\}$ and $t_\gamma := \max\{i : \gamma_i > 0\}$. Here, h is the first non-one entry and t is the last non-zero entry. And let the ‘‘width’’ of γ be $w_\gamma := t_\gamma - h_\gamma$. Note that if $\gamma = (1, \dots, 1, 0, \dots, 0)$, then $w_\gamma = -1$. Otherwise, $w_\gamma \geq 0$.

Theorem 5.2: The policy $\gamma^* := (1, \dots, 1, \gamma_{k^*}^*, 0, \dots, 0)$ is an optimal policy among all policies that neglect timing information; here, $S_0 \in \left[\sum_{i=1}^{k^*-1} S_i, \sum_{i=1}^{k^*} S_i \right)$ and $\gamma_{k^*}^* := \frac{S_0 - \sum_{i=1}^{k^*-1} S_i}{S_{k^*}}$; further, it has minimum width among all other timing-ignorant optimal policies.

The proof of theorem (5.2) follows directly from the two lemmas given below.

Lemma 1: An optimal policy γ must meet the cost constraint with equality. The proof is trivial: pick a γ_i that is not 1, increase it until it is either 1 or until the cost constraint is met with equality; this policy has a higher rate than γ .

Lemma 2: Given a policy γ with width $w_\gamma > 0$, there is a policy λ with width $w_\lambda < w_\gamma$ that performs no worse than γ .

Proof: Define $\lambda := (\lambda_1, \dots, \lambda_k)$ as follows:

$$\lambda_i := \begin{cases} \min\{1, \gamma_{h_\gamma} + \gamma_{t_\gamma} \frac{S_{t_\gamma}}{S_{h_\gamma}}\}, & i = h_\gamma \\ \max\{0, \gamma_{t_\gamma} - (1 - \gamma_{h_\gamma}) \frac{S_{h_\gamma}}{S_{t_\gamma}}\}, & i = t_\gamma \\ \gamma_i, & \text{else} \end{cases}$$

The idea is that we borrow from the ‘‘tail’’ of γ to supplement its ‘‘head’’; we will stop when we have either depleted t_γ to 0 or filled h_γ to 1. We end up in one of two cases (or possibly both, in which both cases are consistent):

Case 1: $\lambda_{h_\gamma} = 1$ and $\lambda_{t_\gamma} = \gamma_{t_\gamma} - (1 - \gamma_{h_\gamma}) \frac{S_{h_\gamma}}{S_{t_\gamma}}$. Note that $h_\lambda > h_\gamma$ and $t_\lambda \leq t_\gamma$; hence $w_\lambda < w_\gamma$ and we have decreased the width. Further, we have improved the rate:

$$\begin{aligned} R_\lambda - R_\gamma &= \sum_{j=1}^k \lambda_j \tilde{C}_j - \sum_{j=1}^k \gamma_j \tilde{C}_j \\ &= \tilde{C}_{h_\gamma} (\lambda_{h_\gamma} - \gamma_{h_\gamma}) + \tilde{C}_{t_\gamma} (\lambda_{t_\gamma} - \gamma_{t_\gamma}) \\ &= \tilde{C}_{h_\gamma} (1 - \gamma_{h_\gamma}) + \tilde{C}_{t_\gamma} \left[-(1 - \gamma_{h_\gamma}) \frac{S_{h_\gamma}}{S_{t_\gamma}} \right] \\ &= (1 - \gamma_{h_\gamma}) S_{h_\gamma} \left(\frac{\tilde{C}_{h_\gamma}}{S_{h_\gamma}} - \frac{\tilde{C}_{t_\gamma}}{S_{t_\gamma}} \right) \\ &\geq 0 \end{aligned}$$

Case 2: $\lambda_{h_\gamma} = \gamma_{h_\gamma} + \gamma_{t_\gamma} \frac{S_{t_\gamma}}{S_{h_\gamma}}$ and $\lambda_{t_\gamma} = 0$. Note that $h_\lambda \geq h_\gamma$ and $t_\lambda < t_\gamma$; hence $w_\lambda < w_\gamma$ and we have decreased the width. Further, we have improved the rate:

$$\begin{aligned} R_\lambda - R_\gamma &= \tilde{C}_{h_\gamma} (\lambda_{h_\gamma} - \gamma_{h_\gamma}) + \tilde{C}_{t_\gamma} (\lambda_{t_\gamma} - \gamma_{t_\gamma}) \\ &= \tilde{C}_{h_\gamma} \left(\gamma_{t_\gamma} \frac{S_{t_\gamma}}{S_{h_\gamma}} \right) + \tilde{C}_{t_\gamma} (-\gamma_{t_\gamma}) \\ &= \gamma_{t_\gamma} S_{t_\gamma} \left(\frac{\tilde{C}_{h_\gamma}}{S_{h_\gamma}} - \frac{\tilde{C}_{t_\gamma}}{S_{t_\gamma}} \right) \\ &\geq 0 \end{aligned}$$

Proof: (of Theorem 5.2) Let γ be any valid policy. We first apply Lemma 1 to γ until it satisfies the cost constraint with equality. Then, we apply Lemma 2 to γ until it has width 0 or -1. Each step in this process results in a policy no worse than γ ; the algorithm terminates in no more than $2k$ steps and always ends at γ^* (there is only one way to distribute S and have minimum width). This algorithm works for any valid γ , so we conclude that no policy can outperform γ^* . ■

Proof: (of Theorem 5.1) Theorem 5.2 provides an scheme that achieves a rate $R = \sum_{j=1}^{k^*} \gamma_j^* \tilde{C}_j$. Then the lower bound is proven:

$$C_{par} \geq \sum_{j=1}^{k^*} \gamma_j^* \tilde{C}_j \geq \sum_{j=1}^{k^*} \gamma_j^* (C_j - \frac{1}{m_j}) \geq R^* - \sum_{i=1}^{k^*} \frac{1}{m_i}$$

The upper bound follows directly from equation (6), since γ^* is a solution to the maximization in that equation:

$$\begin{aligned} C_{par} &\leq \sum_{i=1}^k \frac{1}{m_i} + \max_{(\gamma_i)} \sum_{i=1}^k \gamma_i \tilde{C}_i \\ &= \sum_{i=1}^k \frac{1}{m_i} + \sum_{i=1}^k \gamma_i^* \tilde{C}_i \\ &\leq \sum_{i=1}^k \frac{1}{m_i} + R^* \end{aligned}$$

We note that (just like in series case) for any cost constraint, for large packet sizes, the upper and lower bounds match. A typical plot of the achievable rate v/s the allowed cost per time-slot, with large packet sizes, will look like Fig. 5. ■

VI. A SERIES-PARALLEL NETWORK

In this section, we extend the results to a network that can be broken down as a series and parallel combination of channels such as the network shown in Fig. 2. *Throughout this section, we assume the packet sizes are large over each channel.*

Lemma 6.1: The achievable rate(R) of a network is a concave function of the allowed cost constraint per time-slot(S) over the network.

Proof: Let $R = f(S)$ describe the max. rate as a function of the allowed cost per time-slot, $f : [0, \infty) \rightarrow [0, \infty)$, $f(0) = 0$. Since a rate achievable with allowed cost S is

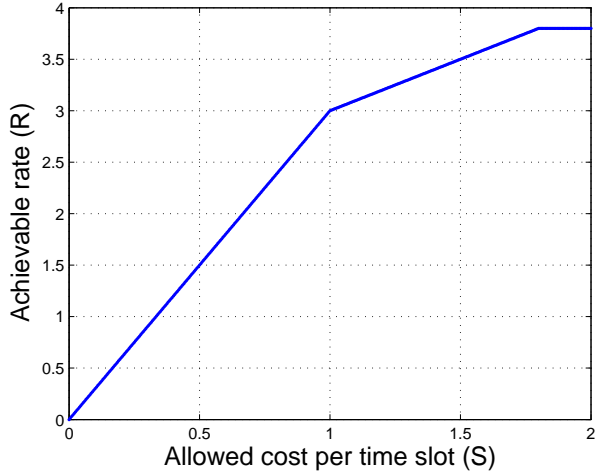


Fig. 5. Parallel Link

achievable with allowed cost greater than S , f is monotonically increasing. We need to prove, for any S_1 and S_2 and $\lambda \in [0, 1]$, that $f(\lambda S_1 + (1 - \lambda)S_2) \geq \lambda f(S_1) + (1 - \lambda)f(S_2)$.

Suppose $R_1 = f(S_1)$. So, the rate R_1 is the maximum achievable rate under the given cost constraint S_1 . Thus, there exists a vector $\Gamma = [\gamma_1, \gamma_2, \dots, \gamma_n]$ that specifies the optimal probabilities for operating each of the n channels in the network. Therefore, if the cost constraint is changed to λS_1 (for $\lambda \in [0, 1]$), a rate λR_1 is achievable; the corresponding operating scheme is given by the vector $\lambda \Gamma$. Thus, $f(\lambda S_1) \geq \lambda f(S_1)$ for $\lambda \in [0, 1]$.

Hence, using the convex combination of the Γ vectors that achieved $f(S_1)$ and $f(S_2)$, we can achieve a rate $\lambda f(S_1) + (1 - \lambda)f(S_2)$, which implies $f(\lambda S_1 + (1 - \lambda)S_2) \geq \lambda f(S_1) + (1 - \lambda)f(S_2)$ and completes the proof. ■

We further prove, through the following two constructive algorithms, that the rate-cost curve of a series-parallel network is always piecewise linear.

Consider the case of two ‘black-boxes’ connected in series. We are given the rate-cost curves for the individual boxes and want to find out the rate-cost curve for the assembly. We assume that the given curves are piecewise linear and prove that the resultant curve is piecewise linear. We propose the following algorithm for constructing the rate-cost curve of the assembly:

1. Superimpose the two curves on top of each other.
2. Given a certain rate R to achieve, each of the series links must be able to sustain a rate of R (since packet-sizes are large, timing plays no role). Therefore, just add the costs the two black-boxes incur in supporting the rate R . Graphically, add the x-co-ordinates of the operating points (S_1, R) and (S_2, R) to get a point on the new rate-cost curve.

(Note: If R is larger than the capacity of either channel, then it’s infeasible.)

Since we are adding the x-co-ordinates of points on two piecewise linear functions, keeping the y-co-ordinates unaltered, we get a piecewise linear curve.

Thus, a series connection of two ‘black boxes’ can be represented as a single equivalent channel. It is relatively straightforward to extend the result to a series assembly of more than two black-boxes.

As an illustration, consider a network that consists of the following two blocks in series:

- 1) A parallel connection of two links,
- 2) A single link.

The (superimposed) rate-allowed cost curves for the two blocks are shown in Fig. 6. The thick curves correspond to the individual blocks in the series combination, the faint line being the maximum possible rate that can be achieved on the series link.

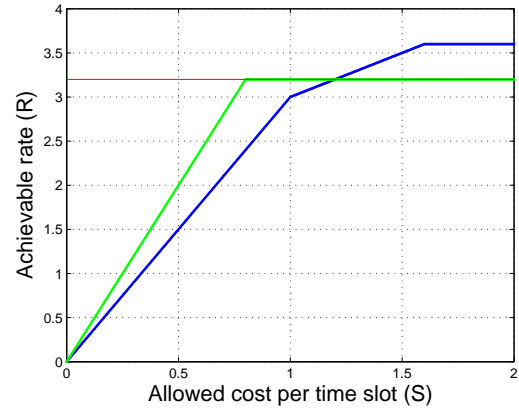


Fig. 6. Parallel-Series link

The rate-allowed cost curve for the equivalent ‘black-box’ is constructed using step 2 of the algorithm, and plotted in Fig. 7.

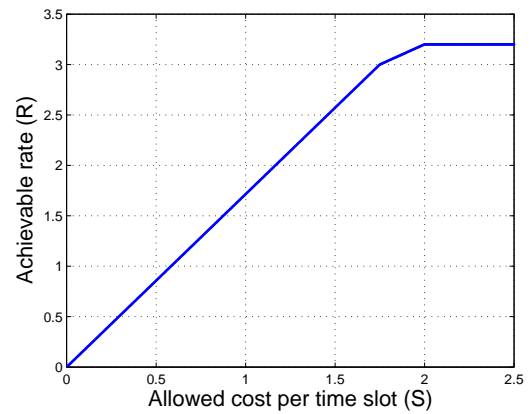


Fig. 7. Equivalent channel

Now, consider a parallel connection of two ‘black-boxes’ whose (piecewise linear, concave) rate-cost curves are spec-

ified. By concavity, each curve is made up of line segments with smaller and smaller slopes until the final (flat) asymptote which indicates the capacity - a rate above which it is impossible to transmit without errors, no matter how much is the allowed cost. The slopes of the segments represent the increment in the achievable rate per unit rise in the allowed cost. The goal is to find the rate-cost curve for the equivalent channel. The following algorithm does the job:

1. Suppose the two curves are made up of line segments $L_{11}, L_{12}, \dots, L_{1k}$ and $L_{21}, L_{22}, \dots, L_{2r}$. Take all these segments and arrange them in the decreasing order of their slopes (ties broken arbitrarily). Let Q_1, Q_2, \dots, Q_{k+r} be this ordering. That is, each Q_i is equal to precisely one $L_{1,j}$ or precisely one $L_{2,j}$, but not both, and $\text{slope}(Q_i) \leq \text{slope}(Q_j)$ if $i > j$.
2. Place Q_1 on the graph (where the final rate-cost curve needs to be plotted) with the lower end at the origin. Place Q_2 with its lower end meeting the top end of Q_1 , and so on. In particular, place Q_{i+1} with its lower end meeting the top end of the segment Q_i . This gives the rate-cost curve for the parallel assembly.

The curve generated by the above algorithm is certainly achievable by construction. For proving its optimality, suppose we have a rate-cost pair (R, S) that can be achieved (with arbitrarily low probability of error) by the parallel assembly and lies above the rate-cost curve generated by the above algorithm. Since the channels are independent and packet sizes are large (making timing information negligible), the rate is the sum of the rates on the two channels, and the cost is the sum of the costs on the individual channels ('black-boxes'). Let (R_1, S_1) and (R_2, S_2) be the two pairs for the channels. Since Q_i s have decreasing slopes, it must be the case that either (R_1, S_1) lies above the rate-cost curve of the first black-box, or (R_2, S_2) lies above the rate-cost curve of the second black-box, or both - a contradiction to the assumption that we can achieve arbitrarily low probability of error over the parallel assembly.

Hence, the rate-cost curve of the parallel assembly of two black-boxes whose individual rate-cost curves are piecewise linear is piecewise linear. It is straightforward to extend this result to a parallel combination of more than two black-boxes.

Since our entire network can be built up by series and parallel assemblies of individual channels (whose rate-cost curves are piecewise linear as proved in section III), we have established that the rate-cost curve for the networks under consideration are piecewise linear. By repeated application of the above two algorithms, we can find the rate-cost curve of the entire network.

VII. SUMMARY

In this paper, we analyze a series parallel network with costs, showing that a flow decomposition is indeed possible. Moreover, a simple algorithm for computing capacity can be found when the packets sizes are large (i.e. when timing capacity is small).

REFERENCES

- [1] J. E. Weiselthier, G. D. Nguyen, and A. Ephremides, "Energy efficient broadcast and multicast trees in wireless networks", in *Mobile Networks and Applications*, vol. 7, 2002, pp. 481492.
- [2] L. Zosin and S. Khuller, "On directed steiner trees", in *In Proc. 13th Annual ACM-SIAM Symp. Discrete Algorithms (SODA 2002)*, San Francisco, CA, Jan 2002.
- [3] A. Ahluwalia, E. Modiano, and L. Shu, "On the complexity and distributed construction of energy efficient broadcast trees in static ad-hoc wireless networks", in *In Proc. 2002 Conf on Information Sci. and Systems (CISS 2002)*, March 2002.
- [4] K. B. Kumar and J. M. Jaffe, "Routing to multiple destinations in computer networks", *IEEE Trans. Commun.*, vol. 31(3), pp. 343351, March 1983.
- [5] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow", *IEEE Tran. Inform. Theory*, vol. 46, no. 4, pp. 12041216, July 2000.
- [6] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "Toward a random operation of networks", Submitted to *IEEE Tran. Inform. Theory*.
- [7] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger, "On randomized network coding", in *In Proc. 41st Annual Allerton Conference on Communication, Control and Computing*, October 2003.
- [8] T. Ho, R. Koetter, M. Medard, D. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting", in *In Proc. 2003 IEEE Intl. Symp. on Inform. Theory (ISIT 2003)*, 2003.
- [9] P. Chou, Y. Wu, and K. Jain, "Practical network coding", in *In Proc. 41st Annual Allerton Conference on Communication, Control and Computing*, October 2003.
- [10] D. Lun, M. Medard, T. Ho, and R. Koetter, "Network coding with a cost criterion", in *In Proc. 2003 IEEE Intl. Symp. on Inform. Theory and its applications (ISITA 2004)*, Oct. 2004.
- [11] D. Lun, N. Ratnakar, R. Koetter, M. Medard, E. Ahmed, and H. Lee, "Achieving minimum-cost multicast : a decentralized approach based on network coding", MIT, LIDS technical report P-2629, Sep. 2004.
- [12] J. Jiless and B. Hajek, "An information-theoretic and game-theoretic study of timing channels", *IEEE Trans. Inform. Theory*, vol. 34, pp. 910917, Sep. 1988.
- [13] V. Anantharam and S. Verdú, "Bits through queues", *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 418, Jan. 1996.
- [14] A. B. Wagner and V. Anantharam, "Information theory of covert timing channels", in *In Proc. of the 2005 NATO/ASI Workshop on Network Security and Intrusion Detection (to appear)*.
- [15] Jain, A. and Chandrasekharan, N., "An efficient parallel algorithm for min-cost flow on directed series-parallel networks", *Proceedings of Seventh International Parallel Processing Symposium*, 1993., 13-16 April 1993 Page(s):188 - 192
- [16] Verdú, S., "On channel capacity per unit cost", *IEEE Transactions on Information Theory*, Volume 36, Issue 5, Sept. 1990 Page(s):1019 - 1030
- [17] R. Koetter, "Flow Decomposition of Capacitated Networks", Talk at ITMANET, Nov 2006.