Correspondence

Network Coding Capacity With a Constrained Number of Coding Nodes

Jillian Cannons, *Student Member, IEEE*, and Kenneth Zeger, *Fellow, IEEE*

***Abstract*—We study network coding capacity under a constraint on the total number of network nodes that can perform coding. That is, only a certain number of network nodes can produce coded outputs, whereas the remaining nodes are limited to performing routing. We prove that every nonnegative, monotonically nondecreasing, eventually constant, rational-valued function on the nonnegative integers is equal to the capacity as a function of the number of allowable coding nodes of some directed acyclic network.**

***Index Terms*—Capacity, flow, information theory, network coding, throughput.**

I. Introduction

Let $\mathbb{N}$ denote the positive integers, and let $\mathbb{R}$ and $\mathbb{Q}$ denote the real and rational numbers, respectively, with a superscript "$+$" denoting restriction to positive values. In this correspondence, a *network* is a directed acyclic multigraph $G = (V, E)$, some of whose nodes are information sources or receivers (e.g., see [13]). Associated with the sources are m generated *messages*, where the ith source message is assumed to be a vector of k_i arbitrary elements of a fixed finite alphabet $\mathcal{A}$ of size at least two. At any node in the network, each out-edge carries a vector of n alphabet symbols which is a function (called an *edge function*) of the vectors of symbols carried on the in-edges to the node, and of the node's message vectors if it is a source. Each network edge is allowed to be used at most once (thus, at most n symbols can travel across each edge). It is assumed that every network edge is reachable by some source message. Associated with each receiver are *demands*, which are subsets of the network messages. Each receiver has *decoding functions* which map the receiver's inputs to vectors of symbols in an attempt to produce the messages demanded at the receiver. The goal is for each receiver to deduce its demanded messages from its in-edges and source messages by having information propagate from the sources through the network.

A $(k_1, \ldots, k_m, n)$ *fractional code* is a collection of edge functions, one for each edge in the network, and decoding functions, one for each demand of each receiver in the network. A $(k_1, \ldots, k_m, n)$ *fractional solution* is a $(k_1, \ldots, k_m, n)$ fractional code which results in every receiver being able to compute its demands via its decoding functions, for all possible assignments of length-k_i vectors over the alphabet to the ith source message, for all i. An edge function performs routing when it copies specified input components to its output components. A node performs *routing* when the edge function of each of its out-edges performs routing. Whenever an edge function for an out-edge of a node depends only on the symbols of a single in-edge of that node, we assume, without loss of generality, that the out-edge carries the same vector of symbols as the in-edge it depends on.

For each i, the ratio k_i/n can be thought of as the rate at which source i injects data into the network. Thus, different sources can produce data at different rates. If a network has a $(k_1, \ldots, k_m, n)$ fractional solution over some alphabet, then we say that $(k_1/n, \ldots, k_m/n)$ is an *achievable rate vector*, and we define the achievable rate region[1] of the network as the set

$$S = \{r \in \mathbb{Q}^m : r \text{ is an achievable rate vector}\}.$$

Determining the achievable rate region of an arbitrary network appears to be a formidable task. Consequently, one typically studies certain scalar quantities called coding capacities, which are related to achievable rates. A routing capacity of a network is a coding capacity under the constraint that only routing is permitted at network nodes. A *coding gain* of a network is the ratio of a coding capacity to a routing capacity. For directed multicast[2] and directed multiple unicast[3] networks, Sanders, Egner, and Tolhuizen [10] and Li and Li [8], respectively, showed that the coding gain can be arbitrarily large.

An important problem is to determine how many nodes in a network are required to perform coding in order for the network to achieve its coding capacity (or to achieve a coding rate arbitrarily close to its capacity if the capacity is not actually achievable). A network node is said to be a *coding node* if at least one of its out-edges has a nonrouting edge function. A similar problem is to determine the number of coding nodes needed to assure the network has a solution (i.e., a $(k_1, \ldots, k_m, n)$ fractional solution with $k_1 = \ldots = k_m = n = 1$). The number of required coding nodes in both problems can in general range anywhere from zero up to the total number of nodes in the network.

For the special case of multicast networks, the problem of finding a minimal set of coding nodes to solve a network has been examined previously in [2], [6], [7], [11]; the results are summarized as follows. Langberg, Sprintson, and Bruck [7] determined upper bounds on the minimum number of coding nodes required for a solution. Their bounds are given as functions of the number of messages and the number of receivers. Tavory, Feder, and Ron [11] showed that with two source messages, the minimum number of coding nodes required for a solution is independent of the total number of nodes in the network, while Fragouli and Soljanin [6] showed this minimum to be upper-bounded by the number of receivers. Bhattad, Ratnakar, Koetter, and Narayanan [2] gave a method for finding solutions with reduced numbers of coding nodes, but their method may not find the minimum possible number of coding nodes. Wu, Jain, and Kung [12] demonstrated that only certain network edges require coding functions. This fact indirectly influences the number of coding nodes required, but does not immediately give an algorithm for finding a minimum node set.

We study here a related (and more general) problem, namely, how network coding capacities can vary as functions of the number of allowable coding nodes. Our main result, given in Theorem III.2, shows

Manuscript received September 26, 2006; revised October 11, 2007. This work was supported by the National Science Foundation, the UCSD Center for Wireless Communications, and the Air Force Office of Scientific Research under Grant FA9550-06-1-0210.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: jcannons@ucsd.edu, zeger@ucsd.edu).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2007.915719

[1]Sometimes in the literature the closure $\bar{S}$, with respect to $\mathbb{R}^m$, is taken as the definition of the *achievable rate region*.

[2]A *multicast* network is a network with a single source and with every receiver demanding all of the source messages.

[3]A *multiple unicast* network is a network where each message is generated by exactly one source node and is demanded by exactly one receiver node.

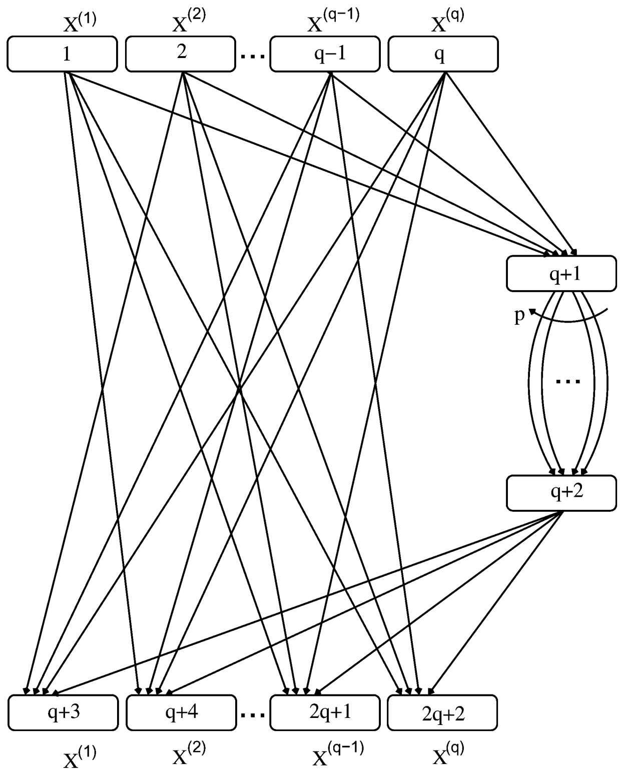

Fig. 1. The network $\mathcal{N}(p,q)$, with $p \leq q$ and $p,q \in \mathbb{Z}^+$. Nodes $n_1,\ldots,n_q$ are the sources, with node n_i providing message $X^{(i)}$, for $1 \leq i \leq q$. Nodes $n_{q+3},\ldots,n_{2q+2}$ are the receivers, with node n_i demanding message $X^{(i-q-2)}$, for $q+3 \leq i \leq 2q+2$. Every source has one out-edge going to node n_{q+1} and every receiver has one in-edge coming from node n_{q+2}. Also, every source n_i has an out-edge going to receiver n_{q+2+j}, for all $j \neq i$. There are p parallel edges from node n_{q+1} to node n_{q+2}.

that the capacities of networks, as functions of the number of allowable coding nodes, can be almost anything. That is, the class of directed acyclic networks can witness arbitrary amounts of coding gain by using arbitrarily sized node subsets for coding.

II. Coding Capacities

Various coding capacities can be defined in terms of the achievable rate region of a network. We study two such quantities, presenting their definitions and determining their values for an example network given in Fig. 1. This network is used to establish Theorem III.2. Li and Li [8] presented a variation of this network and found the routing and coding capacities for the case when $k_i = k$ for all i.

For any $(k_1,\ldots,k_m,n)$ fractional solution, we call the scalar value

$$\frac{1}{m}\left(\frac{k_1}{n}+\ldots+\frac{k_m}{n}\right)$$

an *achievable average rate* of the network. We define the *average coding capacity* of a network to be the supremum of all achievable average rates, namely

$$\mathcal{C}^{\text{average}} = \sup\left\{\frac{1}{m}\sum_{i=1}^{m} r_i : (r_1,\ldots,r_m) \in S\right\}.$$

Similarly, for any $(k_1,\ldots,k_m,n)$ fractional solution, we call the scalar quantity

$$\min\left\{\frac{k_1}{n},\ldots,\frac{k_m}{n}\right\}$$

an *achievable uniform rate* of the network. We define the *uniform coding capacity* of a network to be the supremum of all achievable uniform rates, namely

$$\mathcal{C}^{\text{uniform}} = \sup\left\{\min_{1\leq i\leq m} r_i : (r_1,\ldots,r_m) \in S\right\}.$$

Note that if $r \in S$ and if $r' \in \mathbb{Q}^{m+}$ is component-wise less than or equal to r, then $r' \in S$. In particular, if

$$(r_1,\ldots,r_m) \in S$$

and

$$r_i = \min_{1\leq j\leq m} r_j$$

then

$$(r_i,r_i,\ldots,r_i) \in S$$

which implies

$$\mathcal{C}^{\text{uniform}} = \sup\left\{r_i : (r_1,\ldots,\, r_m) \in S, r_1 = \cdots = r_m\right\}.$$

In other words, all messages can be restricted to having the same dimension

$$k_1 = \cdots = k_m$$

when considering $\mathcal{C}^{\text{uniform}}$.

Also, note that

$$\mathcal{C}^{\text{average}} \geq \mathcal{C}^{\text{uniform}}$$

and that quantities $\mathcal{C}^{\text{average}}$ and $\mathcal{C}^{\text{uniform}}$ are attained by points on the boundary of the closure $\bar{S}$ of S. If a network's edge functions are restricted to purely routing functions, then $\mathcal{C}^{\text{average}}$ and $\mathcal{C}^{\text{uniform}}$ will be referred to as the *average routing capacity* and *uniform routing capacity*, and will be denoted $\mathcal{C}_0^{\text{average}}$ and $\mathcal{C}_0^{\text{uniform}}$, respectively.

Example II.1: In this example, we consider the network in Fig. 1. Note that for each $j = 1,\ldots,q$, every path from source node n_j to receiver node n_{q+2+j} contains the edge $e_{j,q+1}$. Thus, we must have $k_j \leq n$ for all j, and therefore

$$k_1+\cdots+k_q \leq qn$$

so $\mathcal{C}^{\text{average}} \leq 1$.

Furthermore, we can obtain a $(k_1,\ldots,k_q,n)$ fractional coding solution with

$$k_1 = \cdots = k_q = n = 1$$

using routing at all nodes except n_{q+1}, which transmits the $\bmod|\mathcal{A}|$ sum of its inputs on one of its out-edges and nothing on its other $p-1$ out-edges. This solution implies that

$$\mathcal{C}^{\text{average}} \geq 1.$$

Thus, we have $\mathcal{C}^{\text{average}} = 1$.

Clearly

$$\mathcal{C}^{\text{uniform}} \leq \mathcal{C}^{\text{average}} = 1.$$

The presented $(k_1,\ldots,k_q,n)$ fractional coding solution uses

$$k_1 = \cdots = k_q$$

so

$$\mathcal{C}^{\text{uniform}} \geq 1.$$

Thus

$$\mathcal{C}^{\text{uniform}} = 1.$$

When only routing is allowed, all of the messages must pass through the p edges from node n_{q+1} to n_{q+2}. Thus, we must have

$$k_1 + \cdots + k_q \le pn$$

or equivalently

$$\frac{k_1 + \cdots + k_q}{qn} \le \frac{p}{q}.$$

This implies

$$\mathcal{C}_0^{\text{average}} \le \frac{p}{q}.$$

A $(k_1, \ldots, k_q, n)$ fractional routing solution consists of taking

$$k_1 = \cdots = k_q = p$$

and $n = q$ and sending each message $X^{(j)}$ along the corresponding edge $e_{j,q+1}$, sending all

$$k_1 + \cdots + k_q = qp$$

message components from node n_{q+1} to n_{q+2} in an arbitrary fashion, and then sending each message $X^{(j)}$ from node n_{q+2} to the corresponding receiver node n_{q+2+j}. Hence

$$\mathcal{C}_0^{\text{uniform}} \ge \frac{p}{q}$$

and therefore

$$\frac{p}{q} \le \mathcal{C}_0^{\text{uniform}} \le \mathcal{C}_0^{\text{average}} \le \frac{p}{q}.$$

Thus

$$\mathcal{C}_0^{\text{uniform}} = \mathcal{C}_0^{\text{average}} = \frac{p}{q}.$$

Various properties of network routing and coding capacities relating to their relative values, linearity, alphabet size, achievability, and computability have previously been studied [1], [3]–[5], [9]. However, it is not presently known whether or not there exist algorithms that can compute the coding capacity (uniform or average) of an arbitrary network. In fact, computing the exact coding capacity of even relatively simple networks can be a seemingly nontrivial task. At present, very few exact coding capacities have been rigorously derived in the literature.

III. Node-Limited Coding Capacities

For each nonnegative integer i, a $(k_1, \ldots, k_m, n)$ *fractional i-node coding solution* for a network is a $(k_1, \ldots, k_m, n)$ *fractional coding solution* with at most i coding nodes (i.e., having output edges with nonrouting edge functions).[4] For each i, we denote by $\mathcal{C}_i^{\text{average}}$ and $\mathcal{C}_i^{\text{uniform}}$ the average and uniform coding capacities, respectively, when solutions are restricted to those having at most i coding nodes. We make the convention that, for all $i > |V|$

$$\mathcal{C}_i^{\text{average}} = \mathcal{C}_{|V|}^{\text{average}}$$

and

$$\mathcal{C}_i^{\text{uniform}} = \mathcal{C}_{|V|}^{\text{uniform}}.$$

We call $\mathcal{C}_i^{\text{average}}$ and $\mathcal{C}_i^{\text{uniform}}$ the *node-limited average capacity function* and *node-limited uniform capacity function*, respectively. Clearly, the minimum number of coding nodes needed to obtain the average or uniform network capacity is the smallest i such that

$$\mathcal{C}_i^{\text{average}} = \mathcal{C}^{\text{average}}$$

or

$$\mathcal{C}_i^{\text{uniform}} = \mathcal{C}^{\text{uniform}}$$

respectively. Also, the quantities $\mathcal{C}_{|V|}^{\text{uniform}}$ and $\mathcal{C}_{|V|}^{\text{average}}$ are, respectively, the uniform and average coding capacities.

[4]Arbitrary decoding is allowed at receiver nodes and receiver nodes only contribute to the total number of coding nodes in a network if they have out-edges performing coding.

Example III.1: For the network in Fig. 1, since $\mathcal{C}^{\text{average}}$ and $\mathcal{C}^{\text{uniform}}$ are both achieved using only a single coding node (as shown in Example II.1), the node-limited capacities are

$$\mathcal{C}_i^{\text{average}} = \mathcal{C}_i^{\text{uniform}} = \begin{cases} p/q, & \text{for } i = 0 \\ 1, & \text{for } i \ge 1. \end{cases} \tag{1}$$

A function $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ is said to be *eventually constant* if there exists an i such that

$$f(i + j) = f(i)$$

for all $j \in \mathbb{N}$. Thus, the node-limited uniform and average capacity functions are eventually constant. A network's node-limited capacity function is also always nonnegative. For a given number of coding nodes, if a network's node-limited capacity is achievable, then it must be rational, and cannot decrease if more nodes are allowed to perform coding (since one can choose not to use extra nodes for coding). By examining the admissible forms of $\mathcal{C}_i^{\text{average}}$ and $\mathcal{C}_i^{\text{uniform}}$ we gain insight into the possible capacity benefits of performing network coding at a limited number of nodes.

Theorem III.2, whose proof appears after Lemma III.4, demonstrates that node-limited capacities of networks can vary more-or-less arbitrarily as functions of the number of allowable coding nodes. Thus, there cannot exist any useful general upper or lower bounds on the node-limited capacity of an arbitrary network (bounds might exist as functions of the properties of specific networks, however).

Theorem III.2: Every monotonically nondecreasing, eventually constant function $f : \mathbb{N} \cup \{0\} \to \mathbb{Q}^+$ is the node-limited average and uniform capacity function of some directed acyclic network.

Two lemmas are now stated (the proofs are simple and therefore omitted) and are then used to prove Theorem III.2.

Lemma III.3: Let $\mathcal{N}$ be a network with node-limited uniform and average coding capacities $\mathcal{C}_i^{\text{uniform}}$ and $\mathcal{C}_i^{\text{average}}$, respectively, and let p be a positive integer. If every message is replaced at its source node by p new independent messages and every receiver has each message demand replaced by a demand for all of the p new corresponding messages, then the node-limited uniform and average coding capacity functions of the resulting network $\mathcal{N}'$ are $(1/p)\mathcal{C}_i^{\text{uniform}}$ and $(1/p)\mathcal{C}_i^{\text{average}}$, respectively.

Lemma III.4: Let $\mathcal{N}$ be a network with node-limited uniform and average coding capacities $\mathcal{C}_i^{\text{uniform}}$ and $\mathcal{C}_i^{\text{average}}$, respectively, and let q be a positive integer. If every directed edge is replaced by q new parallel directed edges in the same orientation, then the node-limited uniform and average coding capacity functions of the resulting network $\mathcal{N}'$ are $q\mathcal{C}_i^{\text{uniform}}$ and $q\mathcal{C}_i^{\text{average}}$, respectively.

Proof of Theorem III.2 : Suppose $f : \mathbb{N} \cup \{0\} \to \mathbb{Q}^+$ is given by

$$f(i) = \begin{cases} p_i/q_i, & \text{for } 0 \le i < s \\ p_s/q_s, & \text{for } i \ge s \end{cases}$$

where

$$p_0, \ldots, p_s, q_0, \ldots, q_s$$

are positive integers such that

$$\frac{p_0}{q_0} \le \frac{p_1}{q_1} \le \cdots \le \frac{p_s}{q_s}.$$

Define the positive integers

$$b = p_s \cdot \text{lcm}\{q_i : 0 \le i < s\} = \text{lcm}\{p_s q_i : 0 \le i < s\} \in \mathbb{N}$$

$$a_i = \frac{p_i/q_i}{p_s/q_s} \cdot b = \frac{p_i q_s}{p_s q_i} \cdot b \in \mathbb{N}$$

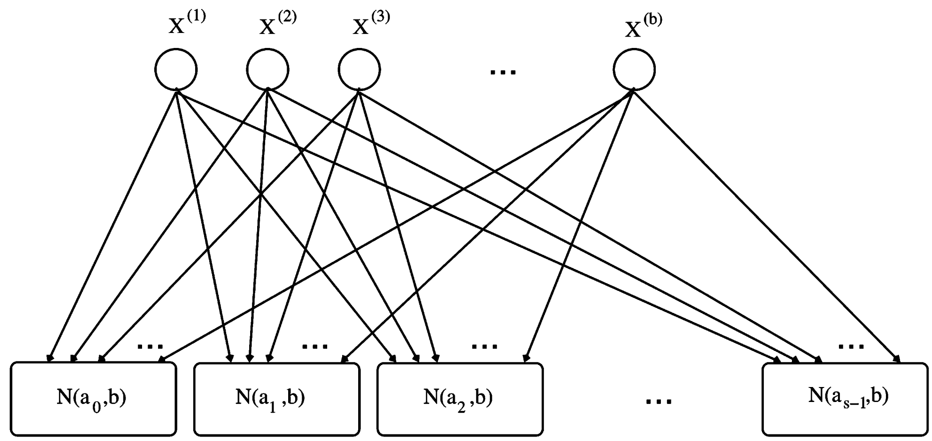

Fig. 2. The network $\mathcal{N}$ has b source nodes, each emitting one message. Each source node has an out-edge to each subblock $\mathcal{N}(a_0, b), \dots, \mathcal{N}(a_{s-1}, b)$. Specifically, in each subblock $\mathcal{N}(a_i, b)$, the previous source messages are removed, however, each previous source node is connected by an in-edge from the unique corresponding source node in $\mathcal{N}$. Each subblock $\mathcal{N}(a_i, b)$ has routing capacity $a_i/b = (p_i/q_i)/(p_s/q_s)$.

and construct a network $\mathcal{N}$ as shown in Fig. 2, which has $m = b$ source messages and uses the networks

$$\mathcal{N}(a_0, b), \dots, \mathcal{N}(a_{s-1}, b)$$

as building blocks (note that $a_i/b \leq 1$ for all i).

Let $\mathcal{C}_i^{\mathrm{uniform}}$ and $\mathcal{C}_i^{\mathrm{average}}$ denote the uniform and average node-limited capacity functions of network $\mathcal{N}$. Also, for $j = 0, \dots, s-1$, let $C_{j,i}^{\mathrm{uniform}}$ and $C_{j,i}^{\mathrm{average}}$ denote the uniform and average node-limited capacity functions of the subblock $\mathcal{N}(a_j, b)$. There are exactly $2s$ nodes in $\mathcal{N}$ that have more than one in-edge and at least one out-edge, and which are therefore potential coding nodes (i.e., two potential coding nodes per subblock). However, for each subblock, any coding performed at the lower potential coding node can be directly incorporated into the upper potential coding node.

For each $i = 0, \dots, s-1$, in order to obtain a $(k_1, \dots, k_m, n)$ fractional i-node coding solution, the quantity

$$\frac{k_1 + \ldots + k_m}{mn}$$

must be at most

$$\min_j \frac{a_j}{b} = \min_j \frac{p_j/q_j}{p_s/q_s}$$

where the minimization is taken over all j for which subblock $\mathcal{N}(a_j, b)$ has no coding nodes (as seen from (1)). That is, we must have

$$\frac{k_1 + \ldots + k_m}{mn} \leq \frac{p_i/q_i}{p_s/q_s}.$$

Therefore, the node-limited average and uniform coding capacities of $\mathcal{N}$ using i coding nodes are at most the respective routing capacities of subblock $\mathcal{N}(a_i, b)$ of $\mathcal{N}$, namely

$$\mathcal{C}_i^{\mathrm{uniform}} \leq C_{i,0}^{\mathrm{uniform}} = \frac{a_i}{b} = \frac{p_i/q_i}{p_s/q_s}$$
$$\mathcal{C}_i^{\mathrm{average}} \leq C_{i,0}^{\mathrm{average}} = \frac{a_i}{b} = \frac{p_i/q_i}{p_s/q_s}.$$

These upper bounds are achievable by using coding at the one useful possible coding node in each of the subblocks

$$\mathcal{N}(a_0, b), \dots, \mathcal{N}(a_{i-1}, b)$$

and using routing elsewhere. By taking

$$\begin{aligned} d &= \mathrm{lcm}(a_i, \dots, a_{s-1}) \\ k_1 = \ldots = k_m &= d \\ n &= bd/a_i \end{aligned}$$

we can obtain a $(k_1, \dots, k_m, n)$ fractional i-node coding solution with coding nodes in subblocks

$$\mathcal{N}(a_0, b), \dots, \mathcal{N}(a_{i-1}, b)$$

and only routing edge-functions in subblocks

$$\mathcal{N}(a_i, b), \dots, \mathcal{N}(a_{s-1}, b).$$

With such a solution, the coding capacity

$$C_{j,1}^{\mathrm{uniform}} = C_{j,1}^{\mathrm{average}} = 1$$

is achieved in each subblock

$$\mathcal{N}(a_0, b), \dots, \mathcal{N}(a_{i-1}, b)$$

and the (unchanging) routing capacity

$$C_{i,0}^{\mathrm{uniform}} = C_{i,0}^{\mathrm{average}}$$

is achieved in each subblock

$$\mathcal{N}(a_i, b), \dots, \mathcal{N}(a_{s-1}, b).$$

Thus, network $\mathcal{N}$ has node-limited average and uniform capacity functions given by

$$\mathcal{C}_i^{\mathrm{average}} = \mathcal{C}_i^{\mathrm{uniform}} = \begin{cases} (p_i/q_i)/(p_s/q_s), & \text{for } 0 \leq i < s \\ 1, & \text{for } i \geq s. \end{cases}$$

By Lemmas III.3 and III.4, if we replace each message of $\mathcal{N}$ by q_s new independent messages and change the receiver demands accordingly, and if we replace each directed edge of $\mathcal{N}$ by p_s parallel edges in the same orientation, then the resulting network $\hat{\mathcal{N}}$ will have node-limited average and uniform capacity functions given by

$$\hat{\mathcal{C}}_i^{\mathrm{average}} = \hat{\mathcal{C}}_i^{\mathrm{uniform}} = \left(\frac{p_s}{q_s}\right) \mathcal{C}_i^{\mathrm{uniform}} = f(i). \qquad \square$$

We note that a simpler network could have been used in the proof of Theorem III.2 if only the case of $\mathcal{C}_i^{\mathrm{uniform}}$ were considered. Namely, we could have used only $\max_{O \leq i < s} q_i p_s$ source nodes and then connected edges from source nodes to subblocks $\mathcal{N}(p_i q_s, q_i p_s)$ as needed.

One consequence of Theorem III.2 is that large coding gains can be suddenly obtained after an arbitrary number of nodes has been used for coding. For example, for any integer $i \geq 0$ and for any real number $t > 0$, there exists a network such that

$$\begin{aligned}\mathcal{C}_0^{\text{uniform}} &= \mathcal{C}_1^{\text{uniform}} = \cdots = \mathcal{C}_i^{\text{uniform}} \\ \mathcal{C}_0^{\text{average}} &= \mathcal{C}_1^{\text{average}} = \cdots = \mathcal{C}_i^{\text{average}} \\ \mathcal{C}_{i+1}^{\text{uniform}} &- \mathcal{C}_i^{\text{uniform}} > t \\ \mathcal{C}_{i+1}^{\text{average}} &- \mathcal{C}_i^{\text{average}} > t.\end{aligned}$$

In Theorem III.2, the existence of networks that achieve prescribed rational-valued node-limited capacity functions was established. It is known in general that not all networks necessarily achieve their capacities [5]. It is presently unknown, however, whether a network coding capacity could be irrational.[5] Thus, we are not presently able to extend Theorem III.2 to real-valued functions. Nevertheless, Theorem III.2 does immediately imply the following asymptotic achievability result for real-valued functions.

Corollary III.5: Every monotonically nondecreasing, eventually constant function $f : \mathbb{N} \cup \{0\} \to \mathbb{R}^+$ is the limit of the node-limited uniform and average capacity function of some sequence of directed acyclic networks.

REFERENCES

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

[2] K. Bhattad, N. Ratnakar, R. Koetter, and K. R. Narayanan, "Minimal network coding for multicast," in *Proc. 2005 IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005.

[3] J. Cannons, R. Dougherty, C. Freiling, and K. Zeger, "Network routing capacity," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 777–788, Mar. 2006.

[4] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.

[5] R. Dougherty, C. Freiling, and K. Zeger, "Unachievability of network coding capacity," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2365–2372, Jun. 2006, Joint issue with *IEEE/ACM Trans. Netw*.

[6] C. Fragouli and E. Soljanin, "Information flow decomposition for network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 829–848, Mar. 2006.

[7] M. Langberg, A. Sprintson, and J. Bruck, "The encoding complexity of network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2386–2397, Jun. 2006, Joint issue with *IEEE/ACM Trans. Netw*.

[8] Z. Li and B. Li, "Network coding: The case of multiple unicast sessions," in *Proc. 42nd Ann. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[9] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[10] P. Sanders, S. Egner, and L. Tolhuizen, "Polynomial time algorithms for network information flow," in *Proc. 15th Ann. ACM Symp. Parallelism in Algorithms and Architectures (SPAA)*, San Diego, CA, Jun. 2003, pp. 286–294.

[11] A. Tavory, M. Feder, and D. Ron, "Bounds on linear codes for network multicast," *Proc. Electronic Colloquium on Computational Complexity (ECCC)*, pp. 1–28, 2003.

[12] Y. Wu, K. Jain, and S.-Y. Kung, "A unification of network coding and tree packing (routing) theorems," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2398–2409, Jun. 2006, Joint issue with *IEEE/ACM Trans. Netw*.

[13] R. W. Yeung, *A First Course in Information Theory*. Amsterdam, The Netherlands: Kluwer, 2002.

[5]It would be interesting to understand whether, for example, a node-limited capacity function of a network could take on some rational and some irrational values, and perhaps achieve some values and not achieve other values. We leave this as an open question.

The Sizes of Optimal q-Ary Codes of Weight Three and Distance Four: A Complete Solution

Yeow Meng Chee, Son Hoang Dau, Alan C. H. Ling, and San Ling

***Abstract*—This correspondence introduces two new constructive techniques to complete the determination of the sizes of optimal q-ary codes of constant weight three and distance four.**

***Index Terms*—Constant-weight codes, large sets with holes, sequences.**

I. INTRODUCTION

The determination of $A_q(n, d, w)$, the size of an optimal q-ary code of length n, distance d, and constant weight w (all terms are defined in the next section), has been the subject of study [1]–[25] due to several important applications requiring nonbinary alphabets, such as coding for bandwidth-efficient channels and design of oligonucleotide sequences for DNA computing. Recently, Chee and Ling [1] introduced an effective technique for constructing optimal constant-weight q-ary codes, which allowed the determination of $A_3(n, 4, 3)$ for all n. For $q > 3$, the value of $A_q(n, 4, 3)$ has also been determined, except when $n \geq q$, $n \equiv 4$ or $5 (\bmod 6)$ [1, Th. 13]. Define the equation shown at the bottom of the next page. The upper bound

$$A_q(n, 4, 3) \leq \min \left\{ U_q(n), \binom{n}{3} \right\} \tag{1}$$

has been established in [1 Th. 12]. In each case where the value of $A_q(n, 4, 3)$ has been determined, it is found to meet this upper bound [1, Ths. 13 and 14].

In this correspondence, we determine $A_q(n, 4, 3)$ completely, showing that it meets the upper bound (1) in all cases. First, we extend the technique of [1] to work with large sets with holes. This allows the determination of $A_q(n, 4, 3)$ when $n \equiv 4 \bmod 6$ and $q \leq n$, or when $n \equiv 5 \bmod 6$ and $q \leq n - 1$. A novel method based on sequences is then used to determine $A_q(n, 4, 3)$ for the remaining cases when $n = q$.

II. DEFINITIONS AND NOTATIONS

The set of integers $\{1, \ldots, n\}$ is denoted by $[n]$. For q a positive integer, we denote the ring $\mathbb{Z}/q\mathbb{Z}$ by $\mathbb{Z}_q$. The set of all nonzero elements of $\mathbb{Z}_q$ is denoted $\mathbb{Z}_q^*$. The ith coordinate of a vector u is denoted by u_i,

Manuscript received September 10, 2007; revised November 11, 2007. The research of Y. M. Chee and S. Ling was supported in part by the Singapore Ministry of Education under Research Grant T206B2204. This work of A. C. H. Ling was done while he was on sabbatical leave at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637616, Singapore.

Y. M. Chee is with the Interactive Digital Media R&D Program Office, Media Development Authority, Singapore 179369, Singapore. He is also with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637616, Singapore, and the Department of Computer Science, School of Computing, National University of Singapore, Singapore 117590, Singapore (e-mail: ymchee@alumni.uwaterloo.ca).

S. H. Dau and S. Ling are with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637616, Singapore (e-mail: lingsan@ntu.edu.sg).

A. C. H. Ling is with the Department of Computer Science, University of Vermont, Burlington, VT 05405 USA (e-mail: aling@emba.uvm.edu).

Communicated by L. M. G. M. Tolhuizen, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2007.915885