Capacity regions for linear and abelian network codes

Terence H. Chan
Department of Computer Science
University of Regina, Canada
terence@cs.uregina.ca

***Abstract*— While linear network codes are proved suboptimal in general multicast scenarios, the loss of throughput due to the use of linear network codes is still unknown. This paper attempts to investigate the loss in throughput by identifying the capacity regions for linear network codes. We prove that the capacity region can be identified by taking intersection of a set of hyperplanes induced by the network and the convex cone closure of the set of all linear representable entropy functions. We also extend the study of network coding capacity region to abelian network codes which contain linear network codes as a subclass. For the case of two multicast sessions, we obtain an inner bound making use of the convex closure of entropy functions which are abelian group representable**

I. INTRODUCTION

Suppose a content provider needs to send a media file to its huge community of subscribers. In the traditional approach, the company will send the file along one (or multiple) Steiner-tree(s) connecting itself to its subscribers. Clearly, intermediate nodes only need to perform simple replication and forward operations in this routing approach.

Despite its simplicity, it does not necessarily utilize available transmission bandwidth efficiently. **Network coding** [1], [8], [9] generalizes the traditional approach by endowing intermediate nodes with additional data-processing capability to "mix/encode" received data before sending the packets. It is well-known that network coding can significantly increase throughput; the gain can even be exponential in certain scenarios.

The significance of network coding is best illustrated by the following example. Consider a simple network in Figure 1 in which mobile users 1 and 2 are to exchange two blocks (b_1 and b_2) with each other. Clearly, mobile users 1 and 2 can respectively send b_1 and b_2 to the base station. In the traditional approach, the base station will first broadcast b_1, say, followed by b_2, requiring two transmissions. However, if network coding is allowed, one can broadcast one block ($b_1 + b_2$) instead. User 1 can then recover b_2 by subtracting b_1 from $b_1 + b_2$, and similar can be done by user 2 to recover b_1. As a result of network coding, the downlink bandwidth efficiency has increased by 100%.

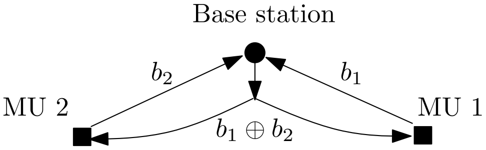

Fig. 1. A simple network: Uplink occurs on separate channels and downlink is a broadcast channel

A fundamental question in network coding is to determine the network coding capacity region (i.e., the capacity of the point-to-point communication channels in order to support a multicast over a network). In the special unicast case (i.e., there is only one data stream to be transmitted to multiple destinations), the capacity region can be determined using the maximal flow / minimum cut bounds. In fact, linear network codes are sufficient to achieve the capacity. In the general case of multicast (i.e., there are more than one independent data streams in the network), network coding capacity region has not been full understood yet. In [5], it was proved that the use of a linear network code is not sufficient to achieve the network coding region. However, the capacity region for linear network codes is still unknown.

The most powerful tool used in characterizing network coding capacity region is by using entropy functions. For example, in [10], inner bounds for network coding region has been obtained by taking the intersection of the set of entropy functions and a collection of hyperplanes induced by the network topology and multicast requirement. This paper uses a similar approach (by using linearly representable entropy functions) to study the network coding capacity region achieved by (timesharing) linear network codes.

A. Entropy vectors

For any set of n random variables $Y_1, \cdots, Y_n$, it induces a set function h such that for any non-empty subset α of $\{1, \cdots, n\}$, $h(\alpha)$ is the joint entropy of $(Y_i : i \in \alpha)$. Clearly, h is non-negative and submodular (i.e., $h(\alpha \cup \beta) + h(\alpha \cap \beta) \leq h(\alpha) + h(\beta)$). We call those functions entropy functions.

Let Γ_n^*[1] [11] be the set of all entropy functions. This set plays an important role in information theory. It has a very complex structure and its full characterization remains an open problem. It was proved in [12] that $\overline{\Gamma}_n^*$, the closure of Γ_n^*, is a closed convex cone. Thus, $\overline{\Gamma}_n^*$ is much more manageable than Γ_n^*, and for many applications, it is sufficient to consider $\overline{\Gamma}_n^*$. Unfortunately, as demonstrated in [10], capacity regions for network codes are obtained by taking the intersection of Γ^* and a set of hyperplanes induced by the underlying network topologies and the multicast requirements. Replacing Γ^* with

---

[1] Γ_n^* depends on n (or the underlying index set for the random variables) in general. If the index set is understood implicitly, we will ignore the dependency to simplify denote it by Γ^*.

its closure leads to an outer bound and it is unknown whether such outer bound is tight or not.

This paper studies the capacity region for linear network codes. Hence, we concern only those entropy vectors that are induced by linear network codes. These entropy vectors are called linearly representable and are formally defined as follows. Unlike the case for general network codes, we will prove in the paper that the capacity region depends only on the closed and convex cone closure of the set of all linearly representable functions. This seems to make the characterization of capacity region for linear network codes more manageable.

Lemma 1: Let $U_1, \cdots, U_n$ be vector subspaces over a finite field $GF(q)$. For any non-empty subset α of $\{1, \cdots, n\}$, let $h(\alpha) = \log q \dim(\oplus_{i\in\alpha} U_i)$. Then h is an entropy vector and is called linearly representable.

Theorem 1: (Ingleton Inequality) Let g be a linearly representable function. Then for any non-empty subset $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ of $\mathcal{N}_n$.

$$\begin{aligned} &g(\alpha_1) + g(\alpha_2) + g(\alpha_3 \cup \alpha_4) \\ &\quad + g(\alpha_1 \cup \alpha_2 \cup \alpha_3) + g(\alpha_1 \cup \alpha_2 \cup \alpha_4) \\ &\qquad\qquad \leq g(\alpha_1 \cup \alpha_2) + (\alpha_1 \cup \alpha_3) + g(\alpha_1 \cup \alpha_4) \\ &\qquad\qquad\quad + g(\alpha_2 \cup \alpha_3) + g(\alpha_2 \cup \alpha_4) \end{aligned}$$

Proof: See [3], [7] ■

Remark: There are entropy vectors that do not satisfy the Ingleton inequality. In the following sections, we show that any linear network codes corresponds to a set of random variables whose entropy vectors are linear representable and satisfy Ingleton inequality as a consequence. Therefore, the existence of entropy vectors that do not satisfy the Ingleton inequality seems to explain why linear network codes are suboptimal.

II. ADMISSIBLE NETWORK CODING REGIONS

We represent a point-to-point communication network by a directed acyclic graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, where the set of nodes V and the set of directed edges E respectively model the set of communication nodes and point-to-point communication channels in the network. For any edges e and f, if the head of f and the tail of e are the same, then we denote the relation by $f \to e$. Similary, if the head of f is a node u, then we denote such a relation by $f \to u$.

For a given communication network G, a multicast requirement M is specified by the following components:

1) an index set S for all independent multicast sessions where each session is a collection of data packets (e.g., by truncating a file) to be multicast to a prescribed set of destination nodes.
2) a mapping $O : \mathsf{S} \to \mathsf{V}$ such that $O(s)$ is the node where data for the sth multicast session is generated;
3) a mapping $D : \mathsf{S} \to 2^{\mathsf{V}}$ such that $D(s)$ is the set of nodes in the network to which data for the sth multicast session is transmitted.

For a given multicast requirement M, the multicast problem is to design a transmission scheme so that data packets for the sth multicast session generated at $O(s)$ can be transmitted to destination nodes in $D(s)$.

A. Network Codes

For a given communication network G and a multicast requirement M, a network code Φ is specified by a set of network coding local functions $\Phi \triangleq \{\phi_e : e \in \mathsf{E}\}$ such that the following criteria are satisfied.

1) For each session s, the input generated in the sth session is a symbol Y_s whose size of support is denoted by $|Y_s|$;
2) For each edge $e \in \mathsf{E}$, let Y_e be the network code symbol transmitted along edge e and the corresponding alphabet size be $|Y_e|$. Then $Y_e = \phi_e(Y_f : f \to e)^2$;
3) For $s \in \mathsf{S}$ and $u \in D(s)$, Y_s can be uniquely determined from $(Y_f : f \to u)$. In other words, $H(Y_s | Y_f : f \to u) = 0$.

In the above network code formulation, there is no restriction on the choice of local network coding functions - both network code symbols and network coding local functions can be arbitrarily selected. However, like designing error control codes for noisy point-to-point channels, network codes with neat algebraic structures are preferred in practice to reduce encoding and decoding complexity. In this paper, we are interested in the characterization of the capacity region of network codes if only special classes of network codes (e.g., linear network codes) are allowed.

Definition 1: A network code $(Y_f : f \in \mathsf{E} \cup \mathsf{S})$ [3] is called linear (over a finite field $GF(q)$) if (1) for any $f \in \mathsf{E} \cup \mathsf{S}$, Y_f is a vector over $GF(q)$, and (2) for any $e \in \mathsf{E}$, Y_e is a linear function of $(Y_f : f \to e)$. In other words, there exists matrices M_e such that

$$Y_e = [Y_f : f \to e] M_e. \tag{1}$$

Remark: In this paper, we always assume that the underlying field is fixed. Therefore, all the linear network codes and the corresponding capacity regions are with respect to the chosen field.

Definition 2: Given a network G and a multicast requirement M, a session-rate link-capacity tuple $(\lambda, \omega) \triangleq (\lambda_s : s \in \mathsf{S}, \omega_e : e \in \mathsf{E})$ is called "(linearly) admissible" if there exists a sequence of (linear) network codes $\Phi^{(n)} = (Y_f^{(n)} : f \in \mathsf{E} \cup \mathsf{S})$ and positive normalizing constants $r(n)$ such that

1) $\forall e \in \mathsf{E}$, $\limsup_{n\to\infty} \log |Y_e^{(n)}| / r(n) \leq \omega_e$,
2) $\forall s \in \mathsf{S}$, $\liminf_{n\to\infty} \log |Y_s^{(n)}| / r(n) \geq \lambda_s$,

Proposition 1: Let $\Phi = (Y_e : e \in \mathsf{E} \cup \mathsf{S})$ be a linear network code with a corresponding session-rate link-capacity

---

[2] To simplify notation, $s \to e$ is used to denote that the sth session is available at the tail of e.

[3] A network code will sometimes be identified by the set of network code symbols $(Y_f : f \in \mathsf{E} \cup \mathsf{S})$ transmitted along all edges and input symbols generated from the sth session.

tuple $(\lambda, \omega) \triangleq (\lambda_s : s \in \mathsf{S},\, \omega_e : e \in \mathsf{E})$. Then there exists vector subspaces $(\mathbf{Y}_f : f \in \mathsf{E} \cup \mathsf{S})$ such that

1) $(\mathbf{Y}_s : s \in \mathsf{S})$ are independent to each other (i.e., $\dim(\oplus_{s:s\in\mathsf{S}}\mathbf{Y}_s) = \sum_{s:s\in\mathsf{S}}\dim(\mathbf{Y}_s)$).
2) for any edge $e \in \mathsf{E}$, $\mathbf{Y}_e$ is a vector subspace of $\oplus_{f:f\to e}\mathbf{Y}_f$ (i.e., $\dim(\oplus_{f:f\to e}\mathbf{Y}_f) = \dim(\oplus_{f:f\to e}\mathbf{Y}_f \oplus \mathbf{Y}_e)$)
3) for $e \in \mathsf{E}$, $\dim(\mathbf{Y}_e) = \omega_e \log_q 2$
4) for $s \in \mathsf{S}$, $\dim(\mathbf{Y}_s \bigcap_{u\in D(s)}(\oplus_{f:f\to u}\mathbf{Y}_f)) = \lambda_s \log_q 2$

Proposition 2: Conversely, if there exists vector subspaces $(\mathbf{Y}_f : f \in \mathsf{E} \cup \mathsf{S})$ such that

1) $(\mathbf{Y}_s : s \in \mathsf{S})$ are independent of each other
2) for any edge $e \in \mathsf{E}$, $\mathbf{Y}_e$ is a vector subspace of $\oplus_{f:f\to e}\mathbf{Y}_f$

Then $(\lambda, \omega) \triangleq (\lambda_s : s \in \mathsf{S},\, \omega_e : e \in \mathsf{E})$ is linearly admissible where $\lambda_s = \dim(\mathbf{Y}_s \bigcap_{u\in D(s)}(\oplus_{f:f\to u}\mathbf{Y}_f))$ and $\omega_e = \dim(\mathbf{Y}_e)$.

B. Capacity region

The previous propositions show that a session-rate link capacity tuple is linearly admissible if and only if there exists vector subspaces $\mathbf{Y}_f$ for $f \in \mathsf{S} \cup \mathsf{E}$ satisfying certain independency and functional dependency relationships. By Lemma 1, the capacity region for linear network codes is then linked directly to those linearly representable entropy functions induced by those vector subspaces.

Intuitively, one can interpret that the rank of $\mathbf{Y}_e$ measures the amount of information transmitted along the edge e. In particular, it means that if two inputs, say $(y_s^1 : s \in \mathsf{S})$ and $(y_s^2 : s \in \mathsf{S})$, are differed by a vector in $\mathbf{Y}_e$, then the node receiving the transmitted symbol on edge e can distinguish the two inputs. As the transmitted symbol on e is a linear function over the inputs, one can "interpret" that $\mathbf{Y}_e$ is the "complement" of the null space of such linear function. Therefore, to study the capacity region of linear network codes, it is sufficient to study vector subspaces that satisfy the independency and functional dependency listed in the two propositions.

Suppose that a network G and a multicast requirement M are given. Let $\mathbb{F}[\mathsf{S},\mathsf{E}]$ be the set of all real-valued functions $\mu(\alpha,\beta)$ where $\alpha \subseteq \mathsf{S}$ and $\beta \subseteq \mathsf{E}$. If $(Y_s : s \in \mathsf{S};\, Y_e : e \in \mathsf{E})$ is a set of random variables, then it induces an element μ in $\mathbb{F}[\mathsf{S},\mathsf{E}]$ such that $\mu(\alpha,\beta)$ is the joint entropy $H(Y_s : s \in \alpha,\, Y_e : e \in \beta)$.

Lemma 2: Let $\Upsilon^*(q)$ be the set of all linearly admissible session-rate link-capacity tuples (with respect to the finite field $GF(q)$). Then $\Upsilon^*(q)$ is a closed and convex cone.

In the following of this section, we will characterize the "capacity region" $\Upsilon^*(q)$. First, we need to define the following subsets of $\mathbb{F}[\mathsf{S},\mathsf{E}]$.

Definition 3: Let $\Delta(q)$ be the subset of $\mathbb{F}[\mathsf{S},\mathsf{E}]$ such that its elements are linearly representable entropy functions (over a finite field $GF(q)$) with respect to random variables $(Y_s : s \in \mathsf{S};\, Y_e : e \in \mathsf{E})$.

Since elements in $\Delta(q)$ are linearly representable, they satisfy the Ingleton inequalities.

Definition 4: For any $e \in \mathsf{E}$, $s \in \mathsf{S}$ and $u \in D(s)$, we define subsets $\mathcal{H}_e, \mathcal{H}_{s,u}$ and $\mathcal{H}_o$ of $\mathbb{F}[\mathsf{S},\mathsf{E}]$ as follows:

$$\begin{aligned}
\mathcal{H}_e &\triangleq \{h : h(\{e, f : f \to e\}) = h(\{f : f \to e\})\} \\
\mathcal{H}_{s,u} &\triangleq \{h : h(s, \{f : f \to u\}) = h(\{f : f \to u\})\} \\
\mathcal{H}_o &\triangleq \{h : h(\mathsf{S}) = \sum_{s\in\mathsf{S}} h(s)\}
\end{aligned}$$

Furthermore, we let

$$\mathcal{H} \triangleq \bigcap_{e\in\mathsf{E}}\mathcal{H}_e \bigcap \mathcal{H}_o \bigcap_{s\in\mathsf{S},u\in D(s)}\mathcal{H}_{s,u}. \tag{2}$$

Theorem 2 (Capacity region): If $h \in \overline{con}(\mathcal{H}\bigcap\Delta(q))^4$, then ν is linearly admissible where for any $s \in \mathsf{S}$ and $e \in \mathsf{E}$, $\nu(s) = h(s)$ and $\nu(e) = h(e)$. Conversely, if ν is linearly admissible, then there exists $h \in \overline{con}(\mathcal{H}\bigcap\Delta(q))$ such that $\nu(s) \le h(s)$ and $\nu(e) \ge h(e)$.

Proof: Since elements in $\Delta(q)$ are linearly representable, it is easy to prove that $\mathcal{H}\bigcap\Delta(q)$ is an inner bound of $\Upsilon^*(q)$ and every linear network code induces a set of random variables whose entropy vector is also in $\mathcal{H}\bigcap\Delta(q)$. The theorem then follows from that $\Upsilon^*(q)$ is a closed and convex cone. ■

Corollary 1 (Outer bound): If a session rate link capacity tuple $(\lambda_s : s \in \mathsf{S},\, \omega_e : e \in \mathsf{E})$ is linearly admissible, then there exists $h \in \mathcal{H}\bigcap\overline{con}(\Delta(q))$ such that $\nu(s) \le h(s)$ and $\nu(e) \ge h(e)$.

To obtain the capacity region for linear network codes, Theorem 2 said the key is to find the intersection of $\mathcal{H}\bigcap\Delta(q)$. While the set $\mathcal{H}$ has an explicit characterization, $\Delta(q)$ is much harder to find. Furthermore, if the network changes, then the intersection of $\mathcal{H}\bigcap\Delta(q)$ will also change.

However, if the outer bound is tight, then the capacity region for linear network codes can be simplified by finding $\overline{con}(\Delta(q))$ (which seems to be more manageable). Then the capacity region can be obtained by taking intersection of the obtained cone with the set $\mathcal{H}$.

In the following of this section, we will show that the outer bound obtained in Corollary 1 is in fact tight. First, we need the following lemmas.

Lemma 3: Let U and V be abelian subgroups of G. Then we have

$$|U||V| = |U \cap V||U \oplus V|. \tag{3}$$

Corollary, if U and V are vector subspaces over a finite field $GF(q)$, then

$$q^{\dim(U)+\dim(V)} = q^{\dim(U\cap V)+\dim(U\oplus V)}, \tag{4}$$

or equivalently, $\dim(U) + \dim(V) = \dim(U \cap V) + \dim(U \oplus V)$.

---

[4] The set $\overline{con}(A)$ is defined as the minimal closed and convex cone containing A.

Lemma 4: Let U_i, V_i be vector subspaces over a finite field $GF(q)$. Then we have

$$\dim(\oplus_i(U_i \cap V_i)) \geq \dim(\oplus_i U_i) + \sum_i (\dim(U_i \cap V_i) - \dim(U_i))$$

and

$$\dim(\bigcap_i (U_i \cap V_i)) \geq \dim(\bigcap_i U_i) + \sum_i (\dim(U_i \cap V_i) - \dim(U_i))$$

Proof:

$$\begin{aligned}
&\dim(U_1 \cap V_1 \cap U_2) \\
&= \dim(U_1 \cap V_1) + \dim(U_2) - \dim((U_1 \cap V_1) \oplus U_2) \\
&\geq \dim(U_1 \cap V_1) + \dim(U_2) - \dim(U_1 \oplus U_2) \\
&= \dim(U_1 \cap U_2) + \dim(U_1 \cap V_1) - \dim(U_1)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\dim((U_1 \cap V_1) \oplus U_2) \\
&= \dim(U_1 \cap V_1) + \dim(U_2) - \dim(U_1 \cap V_1 \cap U_2) \\
&\geq \dim(U_1 \cap V_1) + \dim(U_2) - \dim(U_1 \cap U_2) \\
&= \dim(U_1 \oplus U_2) + \dim(U_1 \cap V_1) - \dim(U_1)
\end{aligned}$$

Lemma 4 can then be proved by induction. ■

This lemma has a simple interpretation. It says that if U_i and $U_i \bigcap V_i$ are almost the same (in the sense of their intersection is almost the same as U_i), then $\oplus_i U_i$ (or $\bigcap_i U_i$) is almost the same as $\oplus_i(U_i \cap V_i)$ (or $\bigcap_i (U_i \cap V_i)$).

Proposition 3: Let $(\mathbf{Y}_s : s \in \mathsf{S}, \mathbf{Y}_e : e \in \mathsf{E})$ be vector subspaces over a finite field $GF(q)$ such that for $s \in \mathsf{S}$, $e \in \mathsf{E}$ and $u \in D(s)$,

$$\begin{aligned}
\dim(\oplus_{f:f\to e}\mathbf{Y}_f \oplus \mathbf{Y}_e) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f) &= \mathcal{O} \\
\sum_{s\in\mathsf{S}} \dim(\mathbf{Y}_s) - \dim(\oplus_{s\in\mathsf{S}}\mathbf{Y}_s) &= \mathcal{O} \\
\dim(\oplus_{f:f\to u}\mathbf{Y}_f \oplus \mathbf{Y}_s) - \dim(\oplus_{f:f\to u}\mathbf{Y}_f) &= \mathcal{O}
\end{aligned}$$

where $\mathcal{O}$ is a term that converges to 0 as $\sum_{s\in\mathsf{S}} \dim(\mathbf{Y}_s)$ goes to infinity after division by $\sum_{s\in\mathsf{S}} \dim(\mathbf{Y}_s)$.

Then one can construct vector subspaces $(\hat{\mathbf{Y}}_s : s \in \mathsf{S}, \hat{\mathbf{Y}}_e : e \in \mathsf{E})$ such that

1) $(\hat{\mathbf{Y}}_s : s \in \mathsf{S})$ are independent of each other
2) for any edge $e \in \mathsf{E}$, $\hat{\mathbf{Y}}_e$ is a vector subspace of $\oplus_{f:f\to e}\hat{\mathbf{Y}}_f$
3) for $s \in \mathsf{S}$ and $e \in \mathsf{E}$, $\dim(\mathbf{Y}_s) \geq \dim(\hat{\mathbf{Y}}_s) \geq \dim(\mathbf{Y}_s) - \mathcal{O}$ and $\dim(\mathbf{Y}_e) \geq \dim(\hat{\mathbf{Y}}_e) \geq \dim(\mathbf{Y}_e) - \mathcal{O}$
4) for $s \in \mathsf{S}$, $\dim(\hat{\mathbf{Y}}_s \bigcap_{u\in D(s)} \oplus_{f:f\to u}\hat{\mathbf{Y}}_f) \geq \dim(\mathbf{Y}_s) - \mathcal{O}$.

Proof: For any given vector subspaces $(\mathbf{Y}_f : f \in \mathsf{E}, \mathbf{Y}_s : s \in \mathsf{S})$, we will construct vector subspaces $(\hat{\mathbf{Y}}_f : f \in \mathsf{E} \cup \mathsf{S})$ as follows.

First, assume without loss of generality that $\mathsf{S} = \{1, 2, \cdots, |\mathsf{S}|\}$. Start from $s = 1$. We define $\hat{\mathbf{Y}}_s = \mathbf{Y}_s$. Now, for any $s > 1$, we define $\hat{\mathbf{Y}}_s$ recursively such that $\hat{\mathbf{Y}}_s$ is a maximal vector subspace of $\mathbf{Y}_s$ and is independent of $\hat{\mathbf{Y}}_{s'}$ for $s' < s - 1$. Clearly, we have

$$\dim(\hat{\mathbf{Y}}_s) = \dim(\oplus_{i\leq s}\mathbf{Y}_i) - \dim(\oplus_{i<s}\mathbf{Y}_i) \tag{5}$$

Then sum the equations for all $s \in \mathsf{S}$ in both sides and we have

$$\begin{aligned}
\sum_{s\in\mathsf{S}} \dim(\hat{\mathbf{Y}}_s) &= \dim(\oplus_{s\in\mathsf{S}}\mathbf{Y}_s) && (6) \\
&= \sum_{s\in\mathsf{S}} \dim(\mathbf{Y}_s) - \mathcal{O} && (7)
\end{aligned}$$

As $\dim(\mathbf{Y}_s) - \dim(\hat{\mathbf{Y}}_s)$ is non-negative, $\dim(\mathbf{Y}_s) - \dim(\hat{\mathbf{Y}}_s) = \mathcal{O}$.

Second, since the underlying network is directed acyclic, we can define $\hat{\mathbf{Y}}_e$ recusively according to the following relation.

$$\hat{\mathbf{Y}}_e = \mathbf{Y}_e \bigcap (\oplus_{f:f\to e}\hat{\mathbf{Y}}_f). \tag{8}$$

By Lemma 4, we have

$$\begin{aligned}
&\dim(\hat{\mathbf{Y}}_e) \\
&= \dim(\mathbf{Y}_e) + \dim(\oplus_{f:f\to e}\hat{\mathbf{Y}}_f) - \dim(\oplus_{f:f\to e}\hat{\mathbf{Y}}_f \oplus \mathbf{Y}_e) \\
&\geq \dim(\mathbf{Y}_e) + \dim(\oplus_{f:f\to e}\hat{\mathbf{Y}}_f) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f \oplus \mathbf{Y}_e) \\
&= \dim(\mathbf{Y}_e) + \dim(\oplus_{f:f\to e}\hat{\mathbf{Y}}_f) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f) \\
&\qquad + \dim(\oplus_{f:f\to e}\mathbf{Y}_f) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f \oplus \mathbf{Y}_e) \\
&= \dim(\mathbf{Y}_e) + \dim(\oplus_{f:f\to e}\hat{\mathbf{Y}}_f) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f) \\
&\qquad - \mathcal{O} \\
&\geq \dim(\mathbf{Y}_e) + \sum_{f:f\to e} (\dim(\hat{\mathbf{Y}}_f) - \dim(\mathbf{Y}_f)) \\
&\geq \dim(\mathbf{Y}_e) - \mathcal{O}
\end{aligned}$$

By mathematical induction, we can then deduce that $\dim(\hat{\mathbf{Y}}_e) \geq \dim(\mathbf{Y}_e) - \mathcal{O}$ for all $e \in \mathsf{E}$.

Third, for any node u, let $\hat{\mathbf{W}}_u = \oplus_{f:f\to u}\hat{\mathbf{Y}}_f$ and $\mathbf{W}_u = \oplus_{f:f\to u}\mathbf{Y}_f$. Then by Lemma 4, we have

$$\dim(\hat{\mathbf{W}}_u) \geq \dim(\mathbf{W}_u) - \mathcal{O}. \tag{9}$$

Consequently,

$$\begin{aligned}
&\dim(\hat{\mathbf{Y}}_s \bigcap_{u\in D(s)} \hat{\mathbf{W}}_u) \\
&\geq \dim(\mathbf{Y}_s \bigcap_{u\in D(s)} \mathbf{W}_u) - \mathcal{O} \\
&\geq \dim(\mathbf{Y}_s) + \sum_{u\in D(s)} (\dim(\mathbf{Y}_s \cap \mathbf{W}_u) - \dim(\mathbf{Y}_s)) - \mathcal{O} \\
&\geq \dim(\mathbf{Y}_s) + \sum_{u\in D(s)} (\dim(\mathbf{W}_u) - \dim(\mathbf{Y}_s \oplus \mathbf{W}_u)) - \mathcal{O} \\
&\geq \dim(\mathbf{Y}_s) - \mathcal{O}
\end{aligned}$$

■

Theorem 3: The outer bound obtained in Corollary 1 is tight. In other words, if $h \in \mathcal{H} \bigcap \overline{con}(\Delta(q))$, then ν is linearly admissible where for any $s \in \mathsf{S}$ and $e \in \mathsf{E}$, $\nu(s) \leq h(s)$ and $\nu(e) \geq h(e)$.

Proof: Suppose $h \in \mathcal{H} \bigcap \overline{con}(\Delta(q))$. Then there exists a sequence of vector subspaces $(\mathbf{Y}_s^k : s \in \mathsf{S}, \mathbf{Y}_e^k : e \in \mathsf{E})$ over a finite field $GF(q)$ and normalizing constants $r(k)$ (which goes to infinity) such that for all $e \in \mathsf{E}$ and $s \in \mathsf{S}$, (1) $\lim_{k\to\infty} \dim(\mathbf{Y}_e^k)/r(k) = h(e)$, (2) $\lim_{k\to\infty} \dim(\mathbf{Y}_s^k)/r(k) = h(s)$, and (3)

$$\begin{aligned} \dim(\oplus_{f:f\to e}\mathbf{Y}_f^k \oplus \mathbf{Y}_e) - \dim(\oplus_{f:f\to e}\mathbf{Y}_f^k) &= \mathcal{O} \\ \sum_{s\in\mathsf{S}} \dim(\mathbf{Y}_s^k) - \dim(\oplus_{s\in\mathsf{S}}\mathbf{Y}_s^k) &= \mathcal{O} \\ \dim(\oplus_{f:f\to u}\mathbf{Y}_f^k \oplus \mathbf{Y}_s) - \dim(\oplus_{f:f\to u}\mathbf{Y}_f^k) &= \mathcal{O} \end{aligned}$$

By Proposition 3, one can construct vector subspaces $(\hat{\mathbf{Y}}_s^k : s \in \mathsf{S}, \hat{\mathbf{Y}}_e^k : e \in \mathsf{E})$ such that

1) $(\hat{\mathbf{Y}}_f^k : f \in \mathsf{S})$ are independent of each other
2) for any edge $e \in \mathsf{E}$, $\hat{\mathbf{Y}}_e^k$ is a vector subspace of $\oplus_{f:f\to e}\hat{\mathbf{Y}}_f^k$
3) for $s \in \mathsf{S}$ and $e \in \mathsf{S}$, $\dim(\hat{\mathbf{Y}}_s^k) \geq \dim(\mathbf{Y}_s^k) - \mathcal{O}$ and $\dim(\mathbf{Y}_e^k) \geq \dim(\hat{\mathbf{Y}}_e^k) \geq \dim(\mathbf{Y}_e^k) - \mathcal{O}$
4) for $s \in \mathsf{S}$, $\dim(\hat{\mathbf{Y}}_s \bigcap_{u\in D(s)} \oplus_{f:f\to u}\hat{\mathbf{Y}}_f) \geq \dim(\mathbf{Y}_s) - \mathcal{O}$.

Then by Proposition 2 and that $\Upsilon^*(q)$ is a closed and convex cone, ν is linearly admissible. ■

For different finite field $GF(q)$, the corresponding set of linearly admissible network coding region can be possibly quite different. From [5], examples are given to illustrate that linear network codes are suboptimal. Specifically, linearly admissible network coding region can be a proper subset of the general network coding region (without restricting the use of linear codes).

In the following section, we consider another class of network codes called abelian network codes, which contains linear network codes (and time-sharing of which) as a subset. As before, an inner bound for the capacity region for abelian network codes can be obtained by taking the intersection of the minimal closed and convex cone containing the set of all "abelian representable" entropy functions and a collection of hyperplanes governed by the network topology and the multicast requirement.

III. GENERALIZATION - ABELIAN NETWORK CODES

Before we give the definition of abelian network codes, we first recall the definition of linear network codes. In a linear network code, the inputs to the networks (i.e., the data generated in the sessions) form a vector space. Then the symbols transmitted along each edge is a linear function of the vector space. Therefore, if two sets of inputs are differed by a vector which is in the null space of the linear function, then the symbol transmitted along the edge corresponding to the two inputs are the same. Moreover, the null space of the linear function of an outgoing edge of a node must contain the intersection of the null spaces of the linear functions of all incoming edges to that node. From this perspective, the definition for abelian network codes is similar to the one for linear network codes in which they share similar properties.

Definition 5: A network code is called an abelian network code if there exists a finite abelian group G and subgroups $(Y_e : e \in \mathsf{S}\cup\mathsf{E})$ such that the following conditions are satisfied:

1) $\prod_{s\in\mathsf{S}} |Y_s| = |G|^{|\mathsf{S}|-1} |\bigcap_{s\in\mathsf{S}} Y_s|$. (This property ensures that the support of the inputs to the network is a Cartesian product. This requirement is similar to that in linear network codes where the inputs are Cartesian product of vectors generated by the sessions.)
2) On each edge $e \in \mathsf{E} \cup \mathsf{S}$, the set of symbols transmitted along the edge e is the set of all cosets of Y_e in G. Specifically, if the input to the network is g, then gY_e is the symbol transmitted along the edge e.
3) for all $e \in \mathsf{E}$, $\bigcap_{f:f\to e} Y_f$ is a subgroup of Y_e. (This property ensures that the functional dependency, i.e., the outgoing message of a node is a function of the incoming messages to the node)
4) encoding at intermediate node works as follows: for any edge e, it receives an index of a coset of Y_f as input where $f \to e$. The symbol (or the index of the left coset of Y_e) to be transmitted along e is the index of the unique coset of Y_e which contains the intersection of all the left cosets of Y_f.

Remark: The role of the subgroups Y_e is similar to that of the null spaces of the linear coding function.

In the above formulation, we have not mentioned how to map data generated by a session to network code symbols. Let X_s be the minimal subgroup containing Y_s and $\bigcap_{f:f\to u} Y_f$ for all $u \in D(s)$. Then data generated by the sth session will be mapped to cosets of Y_s in G such that no more than one cosets of Y_s are contained in a coset of X_s.

Lemma 5: If G is a vector space and all the subgroups are also vector subspaces, then the abelian network code is reduced to a linear network code. In fact, network codes obtained by timesharing multiple linear network codes are also abelian network codes.

Clearly, the admissible network coding region for abelian network codes is not smaller than the linearly admissible network coding region.

Proposition 4: Let G be a finite group and $\{G_1, \cdots, G_n\}$ are its subgroups. Then it induces a set of random variables $\{U, U_1, \cdots, U_n\}$ such that

1) U is a uniform random variable defined over G
2) U_i is a random variable defined on the set of left cosets of G_i in G
3) for any $\alpha \subseteq \{1, \cdots, n\}$, the entropy of $(U_i : i \in \alpha)$ is equal to $\log |G|/|\bigcap_{i\in\alpha} G_i|$.

Furthermore, a random variable U_i is a function of $(U_j : j \in \alpha)$ if and only if G_i contains $\bigcap_{j\in\alpha} G_j$.

Definition 6: Given subgroups $G_1, G_2, \cdots, G_n$ of a group G, let $h \in H_n$ be defined by $h_\alpha = \log \frac{|G|}{|G_\alpha|}$ for all nonempty subset α. Then h is called group representable by $(G, G_1, \cdots, G_n)$ [3]. If G is also abelian, then h is called abelian group representable.

Theorem 4: Let $G_1, \cdots, G_n$ be subgroups of a finite abelian group G. Let $h_\alpha = \log|G| - \log|\bigcap_{i\in\alpha} G_i|$. Then for any non-empty subset $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ of $\mathcal{N}_n$,

$$\begin{aligned} & g(\alpha_1 \cup \alpha_2) + g(\alpha_1 \cup \alpha_3) + g(\alpha_1 \cup \alpha_4) \\ & + g(\alpha_2 \cup \alpha_3) + g(\alpha_2 \cup \alpha_4) \\ \geq \;& \begin{array}{l} g(\alpha_1) + g(\alpha_2) + g(\alpha_3 \cup \alpha_4) \\ + g(\alpha_1 \cup \alpha_2 \cup \alpha_3) + g(\alpha_1 \cup \alpha_2 \cup \alpha_4) \end{array} \end{aligned}$$

Proof: See [3]. ■

Like for linearly representable functions, we define a set $\Delta(ab)$ to be the subset of $\mathbb{F}[\mathsf{S}, \mathsf{E}]$ such that its elements are abelian group representable entropy functions with respect to random variables $(Y_e : e \in \mathsf{E} \cup \mathsf{S})$. By the above theorem, elements in $\Delta(ab)$ also satisfy the Ingleton inequality, which further implies that $\Delta(ab)$ is a proper subset of the set of all entropy functions Γ^*.

Theorem 5 (Inner bound): Suppose there are only two multicast sessions. If $h \in \mathcal{H} \bigcap \overline{con}(\Delta(ab))$, then ν is linearly admissible where for any $s \in \mathsf{S}$ and $e \in \mathsf{E}$, $\nu(s) \leq h(s)$ and $\nu(e) \geq h(e)$.

Proof: The proof is similar to the one in Theorem 3. ■

IV. CONCLUSION

In this paper, we prove that linearly admissible network coding region can be obtained by taking the intersection of the minimal closed and convex cone containing the set of all linear representable entropy functions and a collection of hyperplanes governed by the network topology and the multicast requirement. This capacity region is quite different from the inner bound in [10] for arbitrary network codes which used the set of all entropy functions satisfying a set of function dependencies. Since linear representable entropy functions satisfy Ingleton inequality which is not satisfied by all entropy functions, we believe this is exactly the reason why linear network codes (and even time sharing of which) are suboptimal in general.

REFERENCES

[1] R. Ahlswede, N. Cai, S.Y.R. Li, and R.W. Yeung, "Network information flow," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 1204-1216, 2000.

[2] T.H. Chan and R.W. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. on Inform. Theory*, vol. 48, pp.1992-1995, July 2002.

[3] T.H. Chan, "Aspects of information inequalities and its applications," M.Phil. Thesis, Dept. of Information Engineering, The Chinese University of Hong Kong, September 1998.

[4] P.A. Chou, Y. Wu, and K. Jain, "Practical network coding," *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2003.

[5] R. Dougherty, C. Freiling, and K. Zeger "Insufficiency of Linear Network Codes," *2005 IEEE International Symposium on Information Theory* Adelaide, Australia, pp. 264-267, September 2005.

[6] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "Toward a random operation of networks," submitted to *IEEE Trans. on Inform. Theory* in 2004.

[7] A.W. Ingleton, "Representation of Matroids". Combinatorial mathematics and it's application. Academic Press, 1971.

[8] R. Koetter, and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. on Networking*, vol. 11, pp. 782-795, 2003.

[9] S.Y.R. Li, R.W. Yeung, and N. Cai. "Linear network coding," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 371-381, 2003.

[10] L. Song, R.W. Yeung, N. Cai, "Zero-error network coding for acyclic networks," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 3129-3139, 2003.

[11] R.W. Yeung, "A framework for linear information inequality", IEEE Transactions of Information Theory, vol. 43, pp. 1924-1934, November, 1997.

[12] Z. Zhang and R. W. Yeung, "A non-Shannon type conditional information inequality," IEEE Transactions of Information Theory, vol. 43, pp. 1982-1986, November, 1997.

[13] Z. Zhang and R. W. Yeung, " On the characterization of entropy function via information inequalities" *IEEE Trans. on Inform. Theory,* vol. 44, pp. 1440-1452, Jul 1998.