A Concatenated Network Coding Scheme for Multimedia Transmission

John MacLaren Walsh
Department of Electrical and Computer Engineering
Drexel University
Philadelphia, PA 19104
Email: jwalsh@ece.drexel.edu

Steven Weber
Department of Electrical and Computer Engineering
Drexel University
Philadelphia, PA 19104
Email: sweber@ece.drexel.edu

Abstract — **We discuss a coding technique for transmitting temporally ordered and/or successive refinements encoded multimedia data (such as MPEG video) over a multi-cast network using recently developed breakthrough network coding techniques. Hybridizing network coding philosophy with that of priority encoded transmission and fountain codes, we provide a method which allows a sink to forward to higher OSI model layers (i.e. the source decoder lying in the application layer) the uncoded source data in order of importance as the network coded packets arrive. This technique avoids the delay presently incurred by waiting for the reception of a full rank network coding matrix before decoding the entire block and forwarding it to higher layers.**

I. Introduction and Context

This paper investigates the problem of multicast transmission of temporally ordered and/or successive refinement encoded multimedia data over lossy wired and wireless networks. We are interested in particular in networks that utilize recently developed breakthrough network coding techniques [1, 2, 3] to transmit at or near the multi-cast capacity. A peculiar delay aspect of these network coding techniques is their requirement that the sinks (destinations) must wait until they have received enough linearly independent packets containing different linear combinations of the same transmitted data before they can pass the original (uncoded) data to upper layers. This prevents any of the transmitted data within the uncoded block from being forwarded to higher layers until several encoded packets have been received, rather than as they arrive, causing extra delays in reception. Because of the delay sensitive nature of multimedia transmission, especially in video and teleconferencing applications, as well as different priorities in successive refinements encoded multimedia data, it is desireable to pass on data as it arrives, with the most important part of the data within the original uncoded packet being passed on first. This second context is also especially important when one considers the alternative scenario that the network coded network may time out [2, 3, 4] before a sufficient number of packets with innovative encoding vectors have been received. In this case, despite the fact that several/many useful packets have been received, as it currently stands all of the data in these packets will have to be retransmitted.

In this paper, we suggest a method allowing for the most important part of the uncoded block of source data to be forwarded to higher layers before all of the packets necessary for network decoding of the complete block of source data have been received. To do this, at the transmitter we employ a delay mitigating code, which we construct and which we show is rate optimal in sense that it acheives points on the boundary of an associated capacity region. This is done by treating the network coded channel upon the reception of successive innovative packets as a degraded broadcast channel and determining the rate region of that broadcast channel.

This approach provides a nice hybrid between 1) digital fountain codes and priority encoded transmission, the state of the art in multimedia encoding for practical networks, and 2) recently developed network coding. Fountain codes and priority encoded transmission codes treat the connection between the source and sink as a packet loss and reordering point to point channel abstraction, the effects of which are counteracted with (near Shannon limit performing) erasure codes such as LDPC or RS codes. Network coding does not treat the connection between source and sink with a point to point abstraction, and instead re-encodes the data as it travels through the network. The hybrid between these two philosophies is (fittingly) achieved by exploiting a concatenation of an (inner) network code and an (outer) multi-media erasure/delay mitigating code. The decoder for the concatenated code then allows the most important parts of the block of source data to be passed on accurately before all of the necessary linear combinations packets for full decoding of the entire block have been received. In this manner, we have provided a practical scheme for achieving a vision expressed in [5]:

If a receiver receives only a single global encoding vector, then it can recover the symbols in the most important layer. In general, if a receiver receives global encoding vectors with rank k, then it can recover the symbols in the most important k layers. This is especially useful when broadcasting audio or video data, which can be naturally partitioned layers with different perceptual importance.

It should be pointed out, and is indeed a main point

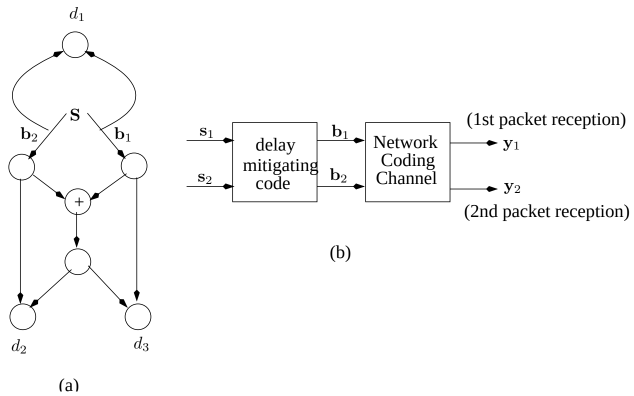

Figure 1: (a): The mantis channel, a slight modification of the butterfly channel. (b): The system proposed in this paper to mitigate delay effects in network coded channels.

of this paper, however, that one may not simply use any PET code unmodified from the routed packet channel to the network coded channel. This subtle point is best explained with the following statement: not all PET codes designed for erasure/loss channels will work as delay codes for network coded channels. Indeed, one must construct a new family of codes (which is necessarily a subfamily of the class of all capable PET codes) to satisfy the new effects of packet decoding blocking and linear combinations that were not present in simple packet reorder and packet loss channels. Fortunately, due to the fact that the family of delay mitigating codes to be concatenated with network codes form a subset of the set of all PET codes, the suggestions in [5] that PET philosophy applies are not too far off mark, as we presently show.

II. Provided Results Overview via Illustrative Example

To provide an overview of our results for the impatient, we first discuss their implications within the context of an appropriate modification of the well known butterfly example [1]. This modification, we which endearingly name the mantis, and is shown in Figure 1 (a) adds an extra destination node above the source node of the classic butterfly. This extra destination node independently receives $\mathbf{b}_1$ and $\mathbf{b}_2$, which are the messages broadcasted from the source node. Let us now endow the butterfly example with an additional notion of packet reception order. In particular, it is entirely unlikely that the pairs of packets arrive at the destination nodes d_1, d_2, d_3 simultaneously, rather at each node one packet will arrive before another. To simplify matters assume for the time being that each of the two orderings of packet receptions are equally likely. Furthermore, suppose the source has some temporal ordering and that blocking delays in the network are unwanted. In particular, we suppose that the source decoder that is to be run at each of the destination nodes can not make use of the second half $\mathbf{s}_2$ of the source data until it has received the first half $\mathbf{s}_1$ but *can* make use of the first half $\mathbf{s}_1$ before obtaining the second half $\mathbf{s}_2$. This can occur, for instance, if successive refinements encoding techniques were used to obtain $\mathbf{s}_1$ and $\mathbf{s}_2$, or if the source encoding technique (as so many digital audio and video formats do) retained the temporal ordering of the original source. From the perspective of this source decoder at each destination, in order to minimize delay, we would like to ensure that $\mathbf{s}_1$ is decodeable from the first packet arrival $\mathbf{y}_1$ and that $\mathbf{s}_2$ is decodeable upon the second innovative packet arrival, i.e. from $\mathbf{y}_1, \mathbf{y}_2$. It is desirable that such ordered decoding occur regardless of the ordering of packet receptions at the destinations for every destination that we could consider.

To allow this to happen, this paper suggests the use of a concatenated coding scheme, as shown in Figure 1 (b), in which a delay mitigating code is incorporated at the source, and then concatenated with the network code. The delay mitigating code sees to it that regardless of which destination node we consider, and regardless of the ordering of packet arrival at that destination node, the first source bit $\mathbf{s}_1$ can be decoded upon receipt of the first packet, and the second source bit $\mathbf{s}_2$ can be decoded upon receipt of the second packet. Of course, the ability to decode the source in order comes at a cost: a decreased overall concatenated code rate relative to a network coded system without the delay mitigating code. Thus, a natural line of inquiry arises which studies the rate delay tradeoffs allowable over all such delay mitigating codes. One way to bound the involved rate delay tradeoffs is to treat the effective source to destination channel (considered over all source destination pairs) as a degraded broadcast channel formed from a linear combinations and erasures channel, in which the receivers receive both received packets and one received packet, respectively. Averaged over all source destination pairs, this becomes equivalent to a degraded broadcast channel depicted in Figure 2 in which the first receiver gets $\mathbf{b}_1$, $\mathbf{b}_2$, or $\mathbf{b}_1 \oplus \mathbf{b}_2$ with equal probability (due to our modification which added the third destination), and the second receiver gets both $\mathbf{b}_1$ and $\mathbf{b}_2$. The rate region of this degraded broadcast channel is shown in Figure 3. In particular, we see that if we do not require any innovative information in the second received packet, then we can receive 1 bit of information in the first received packet, while if we do not require any information in the first received packet, we can receive 2 bits of information in the second received packet. If we wish to receive equal amounts of information upon each of the two packet arrivals (as would be desirable with a constant bit rate source encoder – in the practical sense of those words) then we will transmit at a concatenated code rate of $(\frac{1}{2}+1)^{-1} = 2/3$ of the rate provided by raw network coding.

Although block coding with arbitrarily large block lengths is employed to prove the capacity region of degraded broadcast channels, large length block codes would defeat the purpose of the delay mitigating codes above other than providing bounds, since it would re-

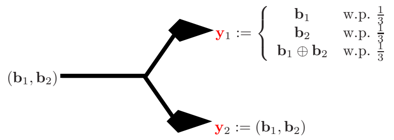

Figure 2: The relevant degraded broadcast channel for studying rate delay tradeoffs in the mantis channel. In this abstracted broadcast channel, the two different receivers correspond to *successive packet arrivals at a single destination* in the network.

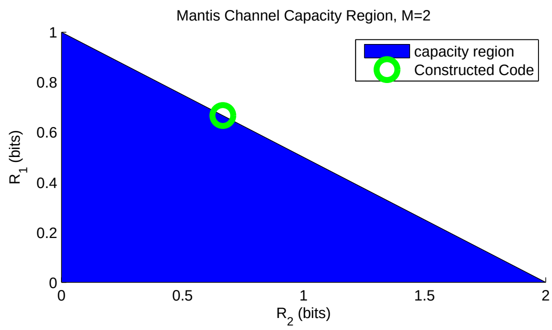

Figure 3: The delay capacity region of the network coded mantis channel.

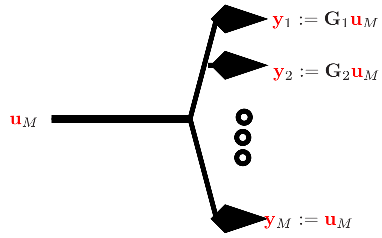

Figure 4: A degraded broadcast channel abstraction of network coded channels used to study rate delay tradeoffs. The different $\mathbf{y}_i$s correspond to the information available to a typical destination node in the network *upon different successive innovative packet arrivals*. Hence, there are as many receivers as the dimension of the global encoding vectors in the network.

quire many uses of the associated broadcast channel. It turns out that one can easily construct simple codes with (degraded broadcast channel use) block length one that achieve all of the rate optimal (max rate for a given delay) points on the capacity region for the mantis. To illustrate this fact, label the source data as $\mathbf{s}_1 = (s_1^1, s_1^2)$ and $\mathbf{s}_2 = (s_2^1, s_2^2)$, with the classic butterfly example taking $\mathbf{b}_1 = \mathbf{s}_1$ and $\mathbf{b}_2 = \mathbf{s}_2$. A simple rate 2/3 delay code, by contrast, is then $\mathbf{b}_1 = (s_1^1, s_1^2, s_2^1)$ and $\mathbf{b}_2 = (s_1^2, s_1^1 \oplus s_1^2, s_2^2)$. A simple check verifies that $\mathbf{s}_1$ can be easily decoded from any of $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_1 \oplus \mathbf{b}_2$, while both $\mathbf{s}_1$ and $\mathbf{s}_2$ can be decoded from both $\mathbf{b}_1$ and $\mathbf{b}_2$ (and thus from any of the pairs $\{\mathbf{b}_1, \mathbf{b}_1 \oplus \mathbf{b}_2\}, \{\mathbf{b}_2, \mathbf{b}_2 \oplus \mathbf{b}_1\}, \{\mathbf{b}_1, \mathbf{b}_2\}$). Thus, regardless of the order that the network coded packets arrive at the destination, and regardless of which destination we consider, we are guaranteed to be able to decode the first source symbol upon the receipt of the first packet, and the second source symbol upon the receipt of the second packet at any destination node. In particular, if the first packet reception is $\mathbf{b}_1$, then s_1^1, s_1^2 are the first two elements of the received packet. Similarly, if the first packet reception is $\mathbf{b}_2$, then s_1^2 is the first element of the received packet, and s_1^1 can be determined by summing the first two bits of the received packet. Finally, if the first recevied packet is $\mathbf{b}_1 \oplus \mathbf{b}_2 = (s_1^1 \oplus s_1^2, s_1^1, s_2^1 \oplus s_2^2)$, then s_1^1 is the second bit in the packet, and s_1^2 is the sum of the first two bits in the packet. One can thus also verify that given both $\mathbf{b}_1$ and $\mathbf{b}_2$ one can determine all of $s_1^1, s_1^2, s_2^1, s_2^2$.

Furthermore, since the point $R_1 = R_2 = \frac{2}{3}$ sits on the boundary of the capacity region associated with mantis channel, we see that this code is also rate optimal (i.e. has maximum overall rate) among all those which deliver equal amounts of decodable information upon the first and second packet receptions. One may verify [6] that codes for the permutation channel, which would correspond to delay mitigating codes for the multi-path routed approach to the mantis channel (and thus only allows reordering of the input packets) have the same rate $\frac{2}{3}$ with respect to uncoded packet for the two packet (multi-path routed) transmission. Thus, we have been able to maintain the rate increase allowed by network coding, but are still able to decode the received packets in order, all through a concatenated source to destination delay mitigating code.

The remainder of the paper sets out to give the arbitary M (dimension of the global encoding vectors) and arbitrary network topology versions of the results just presented for the simple mantis example.

III. Inherent Delay Rate Tradeoffs in Network Coded Channels

From here on, we will consider the random formulation of network coding, reviewed nicely in, e.g. [3], in which innovative global encoding vectors are selected at random as packets proceed through the network. In order to study the rate and delay tradeoffs inherent to network coding, we must first appreciate an abstraction of an effective point to point network coded channel as a degraded broadcast channel. In particular, from the perspective of a representative destination node, we can consider the overall source to destination effect of network coding to be a channel which randomly selects matrices with successively growing rank. In particular, when an innovative packet arrives at the receiver, its encoding vector $\mathbf{g}_k$ is added as a new row in the matrix $\mathbf{G}_k \in GF(q)^{k \times M}$ containing all of the innovative global encoding vectors received thus far, i.e. $\mathbf{G}_k = \left[\mathbf{G}_{k-1}^T, \mathbf{g}_k^T\right]^T$. In these formulations of network coding problems, no data within the original block of data at the source may be decoded at the destination until the encoding matrix has received a sufficient number of rows to become invertible. *Because* *the global encoding vector must be transmitted along with a coded packet, it is desirable for an encoded packet to be very long in order to cut down on the inefficiency of carrying non-information bearing data.* However, this effect exascerbates the decoding delay, for it implies that in the traditional network coding formulation we will have to wait a very long time, indeed, receiving several lengthy packets before we can decode even the earliest or most important bit data from the source. Furthermore, it is possible that the system times out before a sufficient number of innovative packets have been received. In this case, despite the fact that we may have received several innovative packets, all the data within them is wasted and must be retransmitted under the current formulations.

The relevant question then arises: *if we are to consider concatenated delay-mitigating and network coding techniques that will allow us to decode data of successive importance as each network coded packet arrives, what are the rate delay tradeoffs acheivable from the best possible such codes?* Here, when we speak of delay, we are not primarily concerned with packet inter-arrival times, which are indeed studied in [3], but rather in terms of the number of packets which must be received to decode elements of successive importance in the source sequence when utilizing a delay mitigating code between the source coder and the network coding channel. That is, for us, time is slotted on packet arrivals, and we consider the ability to change when these packets arrive as the domain of network coding, while the domain of the delay mitigating code is to allow as much information to be decoded upon each packet arrival as possible. As is usually the case, one answer concerning the fundamental rate and delay tradeoffs can be found for the problem by appealing to (multi-terminal) information theory, much as can be done for priority encoded transmission over packet loss [7, 8] or packet reordering channels [6]. In particular, we formulate the network coded system as a degraded broadcast channel, as shown in Figure 4, in which different receivers correspond to the same destination at different time instants. Here, $\mathbf{B}$ is the matrix of data presented to the network code (output from the delay mitigating code). In particular, the first receiver gets the first innovative output from the channel (i.e. $\mathbf{y}_1 := (\mathbf{g}_1, \mathbf{g}_1\mathbf{B})$, while the second receiver gets the first two innovative packets $\mathbf{y}_2 := (\mathbf{g}_1, \mathbf{g}_1\mathbf{B}, \mathbf{g}_2, \mathbf{g}_2\mathbf{B})$, etc. That this forms a degraded broadcast channel is easily seen from the fact that the received signals at the different receivers are just the signals received at higher elements with positions erased. Since this broadcast channel is degraded its capacity region is known [9], and can be expressed using a Markov chain of dummy message random variables according to $\mathbf{u}_1 \rightarrow \mathbf{u}_2 \rightarrow \cdots \rightarrow \mathbf{u}_M \rightarrow \mathbf{y}_M \rightarrow \mathbf{y}_{M-1} \rightarrow \cdots \rightarrow \mathbf{y}_1$ (where the dummy message random variable $\mathbf{u}_M$ has taken the role of the input matrix of data $\mathbf{B}$ to the network code). This rate region is the closure of the convex hull of the set $\mathcal{R}$ of all rate vectors $\mathbf{R} := (R_1, \ldots, R_M)$ such that $R_1 \leq \mathcal{I}(\mathbf{y}_1; \mathbf{u}_1)$, $R_2 \leq \mathcal{I}(\mathbf{y}_2; \mathbf{u}_2|\mathbf{u}_1)$, $\ldots, R_M \leq \mathcal{I}(\mathbf{y}_M; \mathbf{u}_M|\mathbf{u}_{M-1})$. Here, each R_k is to be interpreted as the amount of new information decodable upon the kth innovative packet arrival for an average destination in the network.

IV. Uniform Network Coded Channel

We consider first the specific case where we can model the source to destination channel created by network coding as sampling the encoding matrices $\mathbf{G}_k$ uniformly from the set $\mathcal{G}_k$ of all encoding matrices in $GF(q)^{k\times M}$ of rank k (henceforth the uniform network coded channel). In this instance, one can bound the capacity region of the abstracted broadcast channel determining the amounts of information decodable upon each successive packet arrival according to the following theorem.

Theorem 1: For the uniform network coded channel

$$\sum_{i=1}^{M} \frac{1}{i} R_i \leq \sum_{\mathbf{G}_1 \in \mathcal{G}_1} \frac{1}{|\mathcal{G}_1|} \mathcal{H}(\mathbf{G}_1 \mathbf{u}_M)$$

Here the right hand side can be bounded by in terms of the number of elements q in the finite field being utilized. Note that in accord with the sugestion from [5], the all rates equal point $R_1 = R_2 = \cdots = R_M$ matches the inverse of the PET girth[8][7], despite the fact that a channel which outputs non-full rank linear combinations is *not* equivalent to an erasure channel, which is the relevant channel for PET. We can prove this theorem with the help of the following proposition

Proposition 1:

$$\frac{k}{k-1} \sum_{\mathbf{G}_{k-1} \in \mathcal{G}_{k-1}} \frac{1}{|\mathcal{G}_{k-1}|} \mathcal{H}(\mathbf{G}_{k-1}\mathbf{u}_M|\mathbf{u}_{k-1}) \geq \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \mathcal{H}(\mathbf{G}_k \mathbf{u}_M|\mathbf{u}_{k-1})$$

To prove the theorem, begin with the equations

$$\begin{aligned} R_1 &\leq \mathcal{I}(\mathbf{y}_1; \mathbf{u}_1) = \mathcal{H}(\mathbf{u}_1) - \sum_{\mathbf{G}_1 \in \mathcal{G}_1} \frac{1}{|\mathcal{G}_1|} \mathcal{H}(\mathbf{u}_1|\mathbf{G}_1\mathbf{u}_M) \\ &= \sum_{\mathbf{G}_1 \in \mathcal{G}_1} \frac{1}{|\mathcal{G}_1|} \left(\mathcal{H}(\mathbf{G}_1\mathbf{u}_M) - \mathcal{H}(\mathbf{G}_1\mathbf{u}_M|\mathbf{u}_1)\right) \end{aligned}$$

Furthermore, for $k \in \{2, \ldots, M\}$, we have

$$\begin{aligned} R_k &\leq \mathcal{I}(\mathbf{y}_k; \mathbf{u}_k|\mathbf{u}_{k-1}) \\ &= \mathcal{H}(\mathbf{u}_k|\mathbf{u}_{k-1}) - \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \mathcal{H}(\mathbf{u}_k|\mathbf{G}_k\mathbf{u}_M, \mathbf{u}_{k-1}) \\ &= \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \left(\mathcal{H}(\mathbf{G}_k\mathbf{u}_M|\mathbf{u}_{k-1}) - \mathcal{H}(\mathbf{G}_k\mathbf{u}_M|\mathbf{u}_k)\right) \end{aligned}$$

Combining the proposition, this bound for $k = 2$, and the bound on R_1, we have

$$\begin{aligned} R_2 &\leq 2\left(\sum_{\mathbf{G}_1 \in \mathcal{G}_1} \frac{1}{|\mathcal{G}_1|} \mathcal{H}(\mathbf{G}_1\mathbf{u}_M) - R_1\right) \\ &\quad - \sum_{\mathbf{G}_2 \in \mathcal{G}_2} \frac{1}{|\mathcal{G}_2|} \mathcal{H}(\mathbf{G}_2\mathbf{u}_M|\mathbf{u}_2) \end{aligned}$$

We may continue repeating this trick until we reach M at which point we have the inequality

$$\sum_{i=1}^{M} \frac{M}{i} R_i \leq M \sum_{\mathbf{G}_1 \in \mathcal{G}_1} \frac{1}{|\mathcal{G}_1|} \mathcal{H}(\mathbf{G}_1 \mathbf{u}_M)$$

which is the theorem we wanted to prove.

To prove the proposition note that we can simply remove any of the k rows from a $\mathbf{G}_k \in \mathcal{G}_k$ to get a valid $\mathbf{G}_{k-1} \in \mathcal{G}_{k-1}$. Furthermore, doing this for each row in every $\mathbf{G}_k \in \mathcal{G}_k$ repeats the same $\mathbf{G}_{k-1}$ $k(q^M - q^{k-1})$ times. Thus,

$$k \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \mathcal{H}(\mathbf{G}_k \mathbf{u}_M | \mathbf{u}_{k-1}) =$$
$$\frac{1}{|\mathcal{G}_k|} \left(\sum_{\mathbf{G}_k \in \mathcal{G}_k} \sum_{i=1}^{k} \mathcal{H}(\mathbf{G}_{k,i} \mathbf{u}_M | \mathbf{G}_{k,\setminus i} \mathbf{u}_M, \mathbf{u}_{k-1}) \right.$$
$$\left. + k(q^M - q^{k-1}) \sum_{\mathbf{G}_{k-1} \in \mathcal{G}_{k-1}} \mathcal{H}(\mathbf{G}_{k-1} \mathbf{u}_M | \mathbf{u}_{k-1}) \right)$$

where $\mathbf{G}_{k,\setminus i}$ is the matrix formed by removing the ith row of $\mathbf{G}_k$ and $\mathbf{G}_{k,i}$ is the ith row. Furthermore because there are q^{k-1} vectors in the span of $\mathbf{G}_{k-1}$ and q^M possible M-vectors, and a new row to be added to $\mathbf{G}_{k-1}$ forms a valid $\mathbf{G}_k$ if it is not in the span of $\mathbf{G}_{k-1}$, we know that $\frac{|\mathcal{G}_k|}{|\mathcal{G}_{k-1}|} = q^M - q^{k-1}$, so we have

$$k \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \mathcal{H}(\mathbf{G}_k \mathbf{u}_M | \mathbf{u}_{k-1}) =$$
$$\frac{1}{|\mathcal{G}_k|} \sum_{\mathbf{G}_k \in \mathcal{G}_k} \sum_{i=1}^{k} \mathcal{H}(\mathbf{G}_{k,i} \mathbf{u}_M | \mathbf{G}_{k,\setminus i} \mathbf{u}_M, \mathbf{u}_{k-1})$$
$$+ k \frac{1}{|\mathcal{G}_{k-1}|} \sum_{\mathbf{G}_{k-1} \in \mathcal{G}_{k-1}} \mathcal{H}(\mathbf{G}_{k-1} \mathbf{u}_M | \mathbf{u}_{k-1})$$

But this then shows via substitution that

$$k \frac{1}{|\mathcal{G}_{k-1}|} \sum_{\mathbf{G}_{k-1} \in \mathcal{G}_{k-1}} \mathcal{H}(\mathbf{G}_{k-1} \mathbf{u}_M | \mathbf{u}_{k-1})$$
$$-(k-1) \sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} \mathcal{H}(\mathbf{G}_k \mathbf{u}_M | \mathbf{u}_{k-1}) =$$
$$\sum_{\mathbf{G}_k \in \mathcal{G}_k} \frac{1}{|\mathcal{G}_k|} (\sum_{i=1}^{k} \mathcal{H}(\mathbf{G}_{k,i} \mathbf{u}_M | \mathbf{G}_{k,1:(i-1)} \mathbf{u}_M, \mathbf{u}_{k-1})$$
$$- \mathcal{H}(\mathbf{G}_{k,i} \mathbf{u}_M | \mathbf{G}_{k,\setminus i} \mathbf{u}_M, \mathbf{u}_{k-1})) \geq 0$$

which verifies the proposition.

We suggest that the bound we have provided is tight by suggesting a construction for capacity achieving delay mitigating codes in Section VI, but first we discuss a method for bounding the general network coded channel capacity region.

V. General Network Coded Channel Delay Capacity Region: Entropy Vectors

Of course, even when an algorithm such as the one discussed in [3] is used to select innovative global encoding vectors randomly, the overall point to point channel (stochastically averaged over destinations) brought about by a multicast network code need not select the full rank encoding matrices uniformly over all such full rank encoding matrices. Indeed, $\mathbf{G}_k$ could be selected according to some other distribution $p_k(\mathbf{G}_k)$ on $\mathcal{G}_k$. In this case, to bound the rate region, which as before describes the amount of information decodable upon successive packet arival instants, we can use a trick similar to bound the rate region of the network code in [10] which, by contrast, described the region of rates simulataneously acheivable to different receivers in a multicast connection. In particular, we first rewrite the delay rate region as

$$R_1 \leq \sum_{\mathbf{G}_1 \in \mathcal{G}_1} p_1(\mathbf{G}_1) \left(\mathcal{H}(\mathbf{G}_1 \mathbf{u}_M) - \mathcal{H}(\mathbf{G}_1 \mathbf{u}_M | \mathbf{u}_1) \right), \ldots$$
$$R_k \leq \sum_{\mathbf{G}_k \in \mathcal{G}_k} p_k(\mathbf{G}_k) \left(\mathcal{H}(\mathbf{G}_k \mathbf{u}_M | \mathbf{u}_{k-1}) - \mathcal{H}(\mathbf{G}_k \mathbf{u}_M | \mathbf{u}_k) \right)$$
$$, \ldots, R_M \leq \mathcal{H}(\mathbf{u}_M | \mathbf{u}_{M-1})$$

At this point the rate region is totally expressed in terms of conditional or joint entropies among dummy message random variables $\mathbf{u}_1, \ldots, \mathbf{u}_M$. We can collect all 2^M joint entropies among these random variables into a vector $\mathbf{h}$ called the entropy function[11, 12]. The set of all possible such $\mathbf{h}$ (over all possible discrete dummy message random variables) is then denoted by $\mathbf{\Gamma}$. Furthermore, the conditional indepence relationships set up by the Markov requirement among the dummy message random variables and the Shannon information inequalities specify linear equalities on $\mathbf{h}$. We can use these information inequalities [11, 12] to analyze this delay rate capacity region as the network coding capacity region was analyzed in [10]. In fact, since we have given our capacity region as a linear map on $\mathbf{h}$, one can use the characterization of $\mathbf{\Gamma}$ from [11, 12] to bound our capacity region using a series of linear programs, much as in [10]. Thus, the reader wishing to provide drawings for a particular non-uniform network coded rate v.s. delay region may do so using the given description of the rate region above and the linear program capacity region bounding technique from [10].

Now that we have bounded the amounts of successive information we can decode at successive packet arrivals, we would like to show that our bounds are tight by providing codes which achieve them.

VI. Capacity Acheiving Delay Mitigating Codes: Uniform Permutation Channel Case

We presently wish to demonstrate that we can construct linear block delay mitigating codes which acheive the all rates equal (i.e. $R_1 = R_2 = \cdots = R_M = \mathsf{H}(M)^{-1}$ with $\mathsf{H}(M)$ proportional to the Mth partial sum of the harmonic series) extremal point of the capacity region. The excellent recent related work [13] shows that the codes discussed are Rank-metric codes, and provides a explicit construction employing Gabidulin codes, which are the analogs of hamming metric based RS codes in the rank metric, and have efficient decoders. Here, due to space constraints, we only discuss the properties of linear codes that will achieve capacity in the language of a partial construction.

In particular, let a rational number representation of this rate $\mathsf{H}(M)^{-1}$ be $\frac{L}{N}$ with N being a desired packet block length for network coding, and L a positive integer multiple of $M!$. Let the block of source encoded data be arranged in successive temporal or importance order as $\mathbf{s} := [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M]^T$ with each source packet being of length L. The output of the delay code is a $M \times N$ matrix $\mathbf{B}$, which is the matrix input to the network code. Let the ith row of $\mathbf{B}$ be denoted by $\mathbf{b}_i$. Break this row up into M parts, $\mathbf{b}_i = \left[\mathbf{b}_i^1, \ldots, \mathbf{b}_i^k, \ldots, \mathbf{b}_i^M\right]$ with the kth part $\mathbf{b}_i^k$ being of length $\frac{L}{k}$ symbols of $GF(q)$ for $k \in \{1, \ldots, M\}$.

Our code constructs $\mathbf{b}_i^k$ using a linear mapping of $\mathbf{s}_k$, so that $\mathbf{b}_i^k := \mathbf{s}_k \mathbf{P}_{i,k}$ here $\mathbf{P}_{i,k} \in GF(q)^{L \times \frac{L}{k}}$. The code construction selects $\{\mathbf{P}_{i,j}\}$ such that the matrix $\mathbf{T}_k \in GF(q)^{L \times L}$ formed from concatenating any k linearly independent M dimensional linear combinations of $\{\mathbf{P}_{i,k} | i \in \{1, \ldots M\}\}$, i.e.

$$\mathbf{T}_k := \left[\sum_{i=1}^{M} c_i^1 \mathbf{P}_{i,k} | \cdots | \sum_{i=1}^{M} c_i^j \mathbf{P}_{i,k} | \cdots | \sum_{i=1}^{M} c_i^k \mathbf{P}_{i,k}\right]$$

is an *invertible* (i.e. full rank) matrix.

Finally, we note that the field size q for the delay mitigating code, which need not exactly match the field size for the network code, must be selected to be large enough so that such a set of generator matrices $\{\mathbf{P}_{i,k}\}$s exist.

Although the codes just constructed to achieve capacity of this rate region are different for the permutation and network coded channels, the bounded rate region we have obtained above is the same as for the permutation channel. Thus, one may use the rate delay tradeoff calculus derived for the permutation channel in [6] to study the relationship between the overall rate of the code and the cumulative delay that a constant bit rate source decoder would experience.

VII. Decoding and Rate Optimality

To see how easily decodeable such a code is, simply note that c_i^j can be taken to be the j, ith element of the encoding matrix $\mathbf{G}_k$ that generates the first through kth innovative recieved packets at a given destination. One then only needs to break the part of $\mathbf{y}_k$ not including the global encoding vectors, which is of dimesion $GF(q)^{k \times N}$, up into subsets of its columns, i.e. $\mathbf{y}_k = \left[\mathbf{y}_k^1 | \cdots | \mathbf{y}_k^j | \ldots | \mathbf{y}_k^M\right]$ with the part $\mathbf{y}_k^j$ having dimension $k \times \frac{L}{j}$. Source $\mathbf{s}_k$ can then be decoded from $\mathbf{y}_k$ by multiplying by the inverse of $\mathbf{T}_k$, i.e. $\mathbf{s}_k = \mathbf{y}_k \mathbf{T}_k^{-1}$. Thus, the code we have specified is rate optimal (by construction), and satisifies the requirement that $\mathbf{s}_1, \ldots, \mathbf{s}_k$ must be decodable from the first k innovative received packets at any destination in the network.

VIII. Conclusions and Future Work

In order to remove the need to wait until enough innovative packets have been received for the encoding matrix to be invertible, this paper has suggested the use of an outer concatenated code to mitigate delay. We have shown limits on the amount of information decodable upon each successive packet arrival at a receiver in a network coded network when using such a delay mitigating code. We have provided an explicit construction of a rate optimal code for mantis channel. [13] provides an efficient construction of the rate optimal codes for other uniform channels. Future work will explore delay coding for non-uniform network coding channels, as well as incorporate inter packet arrival statistics into the analysis and code design.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, p. 12041216, July 2000.

[2] D. S. Lun, M. Médard, and M. Effros, "On coding for reliable communication over packet networks," in *42nd Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2004.

[3] D. S. Lun, "Efficient Operation of Coded Packet Networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[4] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum–cost multicast over coded packet networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2608–2623, June 2006.

[5] P. A. Chou, Y.Wu, and K. Jain, "Practical network coding," in *41st Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2003.

[6] J. M. Walsh and S. Weber, "Coding to reduce delay on a permutation channel," in *45th Allerton Conference on Communication, Control, and Computing*, Sept. 2007.

[7] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority Encoded Transmission," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[8] S. Boucheron and M. R. Slamatian, "About Priority Encoding Transmission," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 699–705, Mar. 2000.

[9] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[10] R. W. Yeung, S. -Y. R. Li, N. Cai, and Z. Zhang, "Network Coding Theory," *Foundation and Trends in Communications and Information Theory*, vol. 2, no. 4-5, pp. 241 –381, 2005.

[11] Z. Zhang and R. W. Yeung, "On Characterization of Entropy Function via Information Inequalities," *IEEE Trans. Inform. Theory*, vol. 44, no. 4, pp. 1440–1452, July 1998.

[12] Z. Zhang and R. W. Yeung, "A Non-Shannon Type Conditional Inequality of Information Quantities," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1982–1986, Nov. 1997.

[13] D. Silva and F. R. Kschischang, "Rank-Metric Codes for Priority Encoding Transmission with Network Coding," in *Canadian Workshop on Information Theory*, Edmonton, AB, Canada, June 2007, p. 8184.