2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, October 2-5, 2018

Latency and Alphabet Size in the Context of Multicast Network Coding

Mira Gonen, Michael Langberg, Alex Sprintson

***Abstract*—We study the relation between *latency* and *alphabet size* in the context of Multicast Network Coding. Given a graph $G = (V, E)$ representing a communication network, a subset $S \subseteq V$ of sources, each of which initially holds a set of information messages, and a set $T \subseteq V$ of terminals; we consider the problem in which one wishes to design a communication scheme that eventually allows all terminals to obtain all the messages held by the sources. In this study we assume that communication is performed in rounds, where in each round each network node may transmit a single (possibly encoded) information packet on any of its outgoing edges. The objective is to minimize the communication latency, i.e., number of communication rounds needed until all terminals have all the messages of the source nodes.**

For sufficiently large alphabet sizes (i.e., large block length, packet sizes), it is known that traditional linear multicast network coding techniques (such as random linear network coding)) minimize latency. In this work we seek to study the task of minimizing latency in the setting of limited alphabet sizes (i.e., finite block length), and alternatively, the task of minimizing the alphabet size in the setting of bounded latency. Through reductive arguments, we prove that it is NP-hard to (i) approximate (and in particular to determine) the minimum alphabet size given a latency constraint; (ii) to approximate (and in particular to determine) the minimum latency of communication schemes in the setting of limited alphabet sizes.

I. Introduction

We consider (a slightly generalized version of) multicast network coding in which given a directed graph $G = (V, E)$ representing a communication network, a subset $S \subseteq V$ of sources, each of which initially holds a (not necessarily disjoint) set of information messages, and a subset of terminals $T \subseteq V$; one wishes to design a multicast communication scheme in which eventually all terminals in T will hold all source messages of the network. Communication is performed in rounds, where in each round every network node may send information on its outgoing edges. In this work, we study the problem of minimizing *communication latency*, i.e., the number of communication rounds needed until all terminals obtain all the messages held by the source nodes.

Several variants of multicast network coding have been studied in the literature (including settings in which the graph G is dynamic and can change over time). The standard objective studied extensively in the literature is that of maximizing the communication rate established between source and terminal nodes, which is characterized by cut-set bounds, e.g., [1]–[4]. In the context of minimizing latency, a number of problem variations have been studied that differ in several aspects including the permitted network operations in each round of communication, the topological assumptions on G, and the source/terminal sets S, T. For example, the setting in which $T = V$ and in each round of communication a single terminal broadcasts a single (possibly encoded) packet to all its neighbors was studied in [5]–[8] when G is a clique and in [9]–[14] for general topologies. This setting is commonly termed "Cooperative Data Exchange". The setting in which G has a general topology, $T = V$, each source in S has a different packet, and in each round of communication each terminal sends/receives (i.e., using the `PULL`, `PUSH`, or `EXCHANGE` model) one information packet on a random outgoing edge (or subset of edges) is usually referred to as the "Gossip" or "Rumor Spreading" problem. This problem has been studied, e.g., in [15]–[23] where upper and lower bounds (that at times depend on topological primitives of the network at hand) on latency (i.e., *stopping time*) are derived. Closely related works also include the works of Wang and Chen and Chekuri et al. [24], [25] that consider the coding advantage in the delay-constrained single-unicast setting.

The results obtained in the studies above are diverse, but a common theme in many of them is the study of random linear network coding as an optimal algorithmic paradigm; i.e., the communication scheme in which each transmitted packet is a linear combination, with uniformly chosen coefficients, of all packets currently present at the transmitter node. For example, Sprintson et al. [6], show that random linear network coding is optimal for the Coded Cooperative Data Exchange problem in the restricted setting of a clique; in a beautiful work by Haeupler [21], tight bounds on the stopping time of algebraic gossip via random linear network coding are derived; and Haeupler et al. in [26] show the optimality of random linear coding under the objective of minimizing latency. The study of [26] focuses on dynamic networks (of general topologies) with limited (and unlimited) buffer sizes. The analysis in the examples above follows the intuition that stems from the study of multicast network coding [1]–[4] in which it is shown that once all terminals in the network want all the information present at source nodes, random linear network coding may achieve optimal results. However, the results above (and in particular that of [26]) assume that communication is done over a sufficiently large field size (i.e., a large alphabet size).

In this work, we study the relations between alphabet size

Mira Gonen is with the Department of Computer Science, Ariel University, Ariel, 40700, Israel (e-mail: mirag@ariel.ac.il).

Michael Langberg is with the Department of Electrical Engineering, State University of New-York at Buffalo, Buffalo, NY 14260, USA (e-mail: mikel@buffalo.edu). Work supported in part by NSF grant 1526771.

Alex Sprintson is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA (e-mail: spalex@tamu.edu). Work supported in part by NSF grants 1642983 and 1718658.

978-1-5386-6596-1/18/$31.00 ©2018 IEEE

and latency. For sufficiently large alphabets, as stated above, latency can be optimized using traditional network coding techniques, such as random linear coding [26]. However, once the alphabet size is restricted, several interesting questions arise in the context of multicast communication. For example, [27]–[30] present multicast network coding instances whose solvability depends strongly on the alphabet size at hand. Closely related to our work is the work of Fragouli and Soljanin [31] which study the computational infeasibility of minimizing the alphabet size while maintaining linear multicast capacity. Although communication latency constraints are not addressed directly by [31], their results can be extended to show the hardness of minimizing latency using linear network coding schemes.

In this work, we consider *general* (i.e., not-necessarily linear) solvability and, roughly speaking, prove that

(i) under fixed sized alphabets, the task of designing optimal, or approximately optimal, latency of communication schemes is intractable;
(ii) under a fixed given latency, the task of designing communication schemes using the optimal, or approximately optimal, alphabet size is intractable.

More specifically, we connect the problems at hand to certain graph and hypergraph coloring problems via reductive arguments. The task of graph coloring is a notoriously difficult one (see, e.g., the survey [32] for formal definitions and results). In particular, it is NP-hard to approximate the chromatic number of a given graph within a factor of $n^{1-\epsilon}$ for any constant $\epsilon > 0$ [33]–[35]. Moreover, the task of determining if a graph is 3-colorable is NP-hard, coloring a 3-colorable graph with 4 colors is NP-hard [36], [37], and coloring a 3-colorable graph with n^{ϵ} colors (for sufficiently small ϵ) is a well known open problem (see e.g. [32], [37] and references therein).

The task of hypergraph coloring is also difficult. There are several generalizations of graph coloring to hypergraphs. We use two variants: *weak coloring* and recent studies on *rainbow coloring*. In weak coloring of hypergraphs the nodes are colored such that no hyperedge is monochromatic. In rainbow coloring of hypergraphs the nodes are colored such that each hyperedge contains at least one node from each color. Guruswami and Lee [38] show that for all constants $k, c \geq 2$ it is NP-hard to distinguish whether a $2k$-uniform hypergraph H is rainbow k-colorable or alternatively if it is not weakly c-colorable. Note that the existence of a rainbow k-coloring for $k \geq 2$ implies that the hypergraph is weakly 2-colorable (and thus weakly c-colorable for any $c \geq 2$). The problem of weakly coloring 2-colorable 3-uniform hypergraphs with any constant number of colors is NP-hard [39], [40].

In what follows we present our model and a detailed description of our results. Our results are proven in Sections III and IV.

II. Model and Statement of Results

Let $G = (V, E)$ be a directed (not necessarily acyclic) network, with $V = [n] = \{1, 2, \ldots, n\}$. Let $S = \{r_1, \ldots r_s\}$ be a set of sources (collectively) holding k messages $X = \{x_1, \ldots x_k\} \in \mathcal{X}^k$ to be delivered to a given subset T of terminals in V. A single source may hold multiple messages, i.e., source r_i holds $X_i \subseteq X$. The messages are elements of a finite alphabet $\mathcal{X}$ of size q and are assumed to be distributed uniformly and independently. Each element in $\mathcal{X}$ is referred to as a *packet*. All non-source nodes do not hold any information at time 0. In each round t of communication, and for each network edge $e = (v, u)$, vertex v may transmit a packet $p_t(v, u)$ in $\mathcal{X}$ to its neighbor u. Packet $p_t(v, u)$ is a function of all information present at v in time $t - 1$ (i.e., all packets v has by round t or messages in X_i if $v = r_i$). For neighbors $u \neq w$, $p_t(v, u)$ is not necessarily equal to $p_t(v, w)$. In this setting, each edge in G has the capacity to send a single information packet in $\mathcal{X}$ in every round. One may model edges of integer capacity by using multiple unit capacity edges. Communication is considered complete when all terminals can decode the messages in X (with zero-error). Note that an edge can transmit multiple packets over multiple communication rounds. This differs from common models of multicast communication (over acylic graphs) in which each edge can only transmit a single packet throughout the communication process.

In our setting, the goal is to allow each terminal in T to decode all messages in X, in such a case we say that all terminals in T are *satisfied*.

Definition 1 (**(q, τ)-feasiblity**) *We say that a communication instance $(G, S, \{X_i\}_{i \in S}, T)$ is* (q, τ)-feasible *if there exists a communication scheme over alphabet $\mathcal{X}$ of size q that satisfies all terminals after τ rounds of communication. We refer to such a scheme as a (q, τ) communication scheme.*

For a given instance $(G, S, \{X_i\}_{i \in S}, T)$, we consider the two following problems:

1) For a given q, design a scheme with $|\mathcal{X}| = q$ that satisfies all terminals in T with the minimum number of communication rounds τ;
2) For a given τ, design a scheme that satisfies all terminals in T after τ rounds of communication using the minimal possible alphabet size $|\mathcal{X}| = q$.

In our analysis, we focus on the *decision* variant of the problems above. Namely, we focus on the problem of deciding whether a given instance $(G, S, \{X_i\}_{i \in S}, T)$ is (q, τ)-feasible or not. Our results address the intractability of finding (or approximating) the optimal q or τ for which $(G, S, \{X_i\}_{i \in S}, T)$ is (q, τ)-feasible. In this context, intractability of the decision problem implies that of explicit code design.

A. Our Results

The following theorems summarize the main results of this work. A number of extended results are given after the theorem proofs in Section III.

Theorem 1 shows the hardness of approximating the minimum required alphabet size q for a given latency constraint τ. Theorem 2 shows the hardness of approximating the latency for a given alphabet size q. Both theorems are formulated on the single-source multicast scenario, showing that even this special case is hard. The claims also hold for the more general case of multicast with multiple sources.

Theorem 1 (Minimizing q given τ) *Let $\tau \geq 2$ be an integer. Then, for a given multicast network coding instance $(G, S, \{X_1\}, T)$ with $|X| = |X_1| = 2$, approximating the minimum q for which $(G, S, \{X_1\}, T)$ is (q, τ)-feasible within any constant multiplicative ratio is NP-hard.*

Theorem 2 (Minimizing τ given q) *Let $k \geq 2$ be an integer. Then, for a given multicast network coding instance $(G, S, \{X_1\}, T)$ with $|X| = |X_1| = k$, and constant q, approximating the minimum τ for which the instance is (q, τ)-feasible within a multiplicative ratio of $k/2$ is NP-hard.*

III. Minimizing alphabet size given a latency constraint

First, as a warm up and to demonstrate our techniques, we show that determining whether an instance $(G, S, \{X_1\}, T)$ is $(2, 2)$ feasible is NP-hard. Next, in Section III-B, we prove the stronger result stated in Theorem 1 that implies that finding an approximate solution within a multiplicative ratio is NP-hard as well.

A. Proof of NP-hardness

We show a reduction is from the `3-Coloring` problem that uses the construction proposed by Fragouli and Soljanin [31, Chapter 7.1.3]. This construction was used in the context of minimizing the alphabet size while achieving linear multicast capacity. In contrast, in this paper we consider *general* (not necessarily linear) encoding functions in the context of minimizing communication latency.

Let $H = (V_H, E_H)$ be a given undirected instance to the `3-Coloring` problem. We construct a directed graph G as follows. The vertex set of G includes a single source node r, $|V_H|$ intermediate vertices v (one vertex for every vertex in H), and $|E_H|$ vertices $t_{(v,u)}$ (one vertex for every edge $(v, u) \in E_H$). The edge set of G includes an edge (r, v) for all $v \in G$ corresponding to vertices in V_H and edges $(v, t_{v,u})$ and $(u, t_{v,u})$ for all vertices $t_{(v,u)}$ in G. The source vertex r holds 2 messages x_1 and x_2 required at the terminal vertices $\{t_{(v,u)}\}_{(v,u)\in E_H}$. Therefore $S = \{r\}$, $X = X_1 = \{x_1, x_2\}$, $T = \{t_{(v,u)} | (v, u) \in E_H\}$. We now show that determining whether $(G, S, \{X_1\}, T)$ is $(2, 2)$ feasible is equivalent to determining if H is 3-colorable.

Lemma 1 *H is 3-colorable if and only if $(G, S, \{X_1\}, T)$ is $(2, 2)$-feasible.*

Proof: Consider a 3-coloring of H. We construct a 2-round communication scheme for G that uses a linear coding scheme over the binary alphabet. In the first round, r transmits, on each of its outgoing edges (r, v), a function of x_1 and x_2 that depends on a color of v. Specifically, if v is colored by the first color in H, r will transmit packet x_1 over edge (r, v). If v is colored by the second or third color, r will transmit x_2 or $x_1 + x_2$ respectively. Here, and in what follows, we use the standard addition over the binary field. In the second round of communication, intermediate vertices v corresponding to vertices in V_H will forward their incoming packets to the corresponding terminal vertices they are connected to. Any terminal vertex $t_{(v,u)}$ corresponding to edge (v, u) in H is guaranteed to receive two different packets on its incoming edges as nodes v and u in H are colored by different colors. This allows $t_{(v,u)}$ to decode both x_1 and x_2.

For the other direction, assume that the instance $(G, S, \{X_1\}, T)$ is $(2, 2)$-feasible. Therefore there exists a two-round communication scheme over the binary alphabet that allows all terminals of G to decode x_1 and x_2. We may assume without loss of generality that vertex r sends some function $f_v(x_1, x_2)$ to intermediate vertex v in round 1, and then every intermediate vertex v forwards $f_v(x_1, x_2)$ to the corresponding terminals it is connected to. The function f_v corresponding to an intermediate vertex v is one out of the six possible Boolean functions over two variables for which $H(f_v(x_1, x_2)) = 1$ (the latter requirement is essential for decoding at terminal nodes). Here, H is the binary entropy function. Note that such functions have an equal number of zeros and ones in their image (going over all input combinations).

The six Boolean functions that satisfy this condition are $x_1, x_2, x_1 + x_2, 1 + x_1, 1 + x_2, 1 + x_1 + x_2$. Here, '+' denotes standard addition over the binary field and '1' correspond to the multiplicative identity. To allow decoding, it holds that terminal nodes cannot be connected to two intermediate vertices that hold one of the pairs $x_1, 1 + x_1$; $x_2, 1 + x_2$; or $x_1 + x_2, 1 + x_1 + x_2$. Thus, one can modify the function $f_v(x_1, x_2) = 1 + x_1$ to be $f_v(x_1, x_2) = x_1$; function $f_v(x_1, x_2) = 1 + x_2$ to be $f_v(x_1, x_2) = x_2$; and function $f_v(x_1, x_2) = 1 + x_1 + x_2$ to be $f_v(x_1, x_2) = x_1 + x_2$; while preserving the ability of all terminals to decode. This implies a 3-coloring of the intermediate nodes of G (one color for each distinct function f_v) in which vertices v, u corresponding to edges (v, u) in H are colored by different colors. ■

We note that the same reduction can be used to prove that computing the scalar linear multicast capacity in network coding instances in which communication is done over the binary field is NP-hard [31] (although over sufficiently large fields computing the scalar linear capacity is polynomial [1]–[4]).

B. Proof of Theorem 1

We now prove Theorem 1, i.e., we show that given τ, approximating the minimum q for which an instance $(G, S, \{X_1\}, T)$ is (q, τ) feasible within any constant multiplicative ratio is NP-hard. Note that for linear functions, the results of [31] imply that it is NP-hard to approximate the minimum q for which the instance $(G, S, \{X_1\}, T)$ is $(q, 2)$ feasible within a multiplicative ratio of $|V|^{1-\epsilon}$ for any constant $\epsilon > 0$. Here, V is the vertex set of G in $(G, S, \{X_1\}, T)$. The novelty in the result below is in the study of general encoding schemes that may use non-linear codes. Our proof uses the fact that for all integers $c, k \geq 2$, given a $2k$-uniform hypergraph H, it is NP-hard to distinguish whether H is rainbow k-colorable or alternatively if it is not even weakly c-colorable [38].

Proof of Theorem 1. For clarity, we assume that $\tau = 2$, the proof can be generalized for any $\tau \geq 2$ using a slight modification of our reduction in which edges are replaced

by (disjoint) paths. Let d be any constant integer. Given a 4-uniform hypergraph $H = (V_H, E_H)$ which is a hard instance of Corollary 1.1 of [38] for $k = 2$ and $c = d^{d^2}$, we construct an instance $(G, S, \{X_1\}, T)$ following the same lines as in the proof of Lemma 1. We show the approximating the minimum q for which $(G, S, \{X_1\}, T)$ is $(q, 2)$ feasible will allow us to determine whether H is rainbow $k = 2$-colorable or alternatively if it is not even weakly $c = d^{d^2}$-colorable [38].

The vertex set of G includes a single source node r, $|V_H|$ intermediate vertices v (one vertex for every vertex in H), and $|E_H|$ vertices t_e (one vertex for every hyperedge $e \in E_H$). The edge set of G includes an edge (r, v) for all $v \in G$ corresponding to vertices in V_H and edges (v, t_e) for all $e \in E_H$, and $v \in e$. Note that here, edge e is a subset of vertices of size 4. The source vertex r holds 2 messages x_1, x_2 required at the terminal vertices $\{t_e\}_{e \in E_H}$. Therefore $S = \{r\}$, $X = X_1 = \{x_1, x_2\}$, $T = \{t_e | e \in E_H\}$.

Consider a 2 rainbow coloring of H. We construct a 2-round communication scheme for G, over an alphabet of size 2, as in the proof of Lemma 1. In the first round, r transmits on its outgoing edges some function of x_1, x_2. For $i \in \{1, 2\}$, for vertex v of color i in H, r will transmit x_i over edge (r, v). In the second round of communication, intermediate vertices v corresponding to vertices in V_H will forward their incoming message to the corresponding terminal vertices they are connected to. Terminal vertex t_e corresponding to edge e in H is guaranteed to receive 2 different packets on its incoming edges as the nodes of e are colored by 2 different colors in H (by its rainbow coloring). This allows t_e to decode.

Assume alternatively that the instance $(G, S, \{X_1\}, T)$ is $(q, 2)$-feasible. Then, there exists a 2-round communication scheme over an alphabet of size q that allows all terminals of G to decode x_1 and x_2. We show that H is c colorable, for $c = q^{q^2}$. We assume without loss of generality that vertex r sends some function $f_v(x_1, x_2)$ to intermediate vertex v in the first round, and that every intermediate vertex v forwards $f_v(x_1, x_2)$ to the corresponding terminals it is connected to in the second round. Notice that since there are 2 rounds, every terminal vertex receives one packet. We define a distinct color for each function. The total number of colors is at most $c = q^{q^2}$ as this is the total number of Boolean functions f_v possible over an alphabet of size q. Since each terminal vertex t_e must decode all packets and each node v with the same color forwards at most one encoded packet to t_e, then it cannot be the case that all vertices in e are colored by the same color (e cannot be monochromatic). Thus we get a weak coloring of H using at most c colors.

Now, for our parameters d and $c = d^{d^2}$, given H which is taken to be either rainbow 2-colorable or alternatively not weakly c-colorable, we can distinguish between the two cases by approximating the minimum q for which $(G, S, \{X_1\}, T)$ is $(q, 2)$-feasible. Namely, if we can approximate the minimum q within a multiplicative ratio α for which $2\alpha = d$ we can distinguish the two cases of H using the following argument. Given H, we construct $(G, S, \{X_1\}, T)$ and approximate the minimum q for which $(G, S, \{X_1\}, T)$ is $(q, 2)$-feasible. Say our approximate value is q_0, implying that we have established that $(G, S, \{X_1\}, T)$ is $(q_0, 2)$-feasible. We now show that if $q_0 \leq d$ it must be that H was originally rainbow 2-colorable, otherwise H was originally not weakly c colorable.

Specifically, if $q_0 \leq d$ then, using our analysis above, H is weakly $q_0^{q_0^2}$-colorable. However, $q_0^{q_0^2} \leq d^{d^2} = c$, which implies that H is weakly c-colorable. Which in turn implies that H must have originally been rainbow 2-colorable. If $q_0 > d$, consider in contradiction that H was originally rainbow 2-colorable. In this case, $(G, S, \{X_1\}, T)$ is $(2, 2)$-feasible, i.e., the minimum q in this case is 2. Thus, $q_0 \leq 2\alpha = d$, resulting in a contradiction.

We conclude that approximating the minimum q for which $(G, S, \{X_1\}, T)$ is $(q, 2)$ feasible within an approximation ratio of $\alpha = d/2$ is NP-hard. As d was chosen arbitrarily, the theorem follows. ■

C. Intractability of non-constant approximation

We now address the question of approximating the minimal alphabet size q within any approximation ratio (not necessarily constant). We present a reduction from the task of coloring graphs H that are known to have $\chi(H) = 3$. Recall that the task of coloring a 3-colorable graph with n^ϵ colors (for sufficiently small ϵ) is a well known open problem, see, e.g., [32], [37]).

Lemma 2 *Let $(G, S, \{X_1\}, T)$ be a multicast network coding instance with $|X| = |X_1| = 2$ which is $(2, 2)$-feasible. Then, for any $q \geq 2$, finding a $(q, 2)$ solution to the instance $(G, S, \{X_1\}, T)$ is as hard as coloring a 3-colorable graph with $q^2 - q + 2$ colors.*

Proof: We use the reduction of Lemma 1. Consider a 3-coloring of H. Then the corresponding $(G, S, \{X_1\}, T)$ is (2,2)-feasible by Lemma 1.

Assume we find a $(q, 2)$-feasible solution for $(G, S, \{X_1\}, T)$. Then there exists a 2-round communication scheme over an alphabet $\mathcal{X}$ of size q that allows all terminals of G to decode x_1 and x_2. We show that this implies a coloring of H with $q^2 - q + 1$ colors. As in the proof of Lemma 1, we may assume without loss of generality that vertex r sends some function $f_v(x_1, x_2)$ to intermediate vertex v in round 1, and then every intermediate vertex v forwards $f_v(x_1, x_2)$ to the corresponding terminals it is connected to. In addition, as before, we may assume that the functions f_v satisfy $H(f_v(x_1, x_2)) = \log q$, as otherwise the terminals could not decode. We now show explicitly that our decoding requirement implies that each f_v is *balanced*, namely that any element in $\mathcal{X}$ has exactly q sources in $\mathcal{X}^2$. That is, for any $a \in \mathcal{X}$, $|\{(x_1, x_2) \mid f_v(x_1, x_2) = a\}| = q$.

First, assume to the contrary that there exists a function f_v used by the given communication scheme for which there exist at least $q + 1$ elements $(x_1^1, x_2^1), (x_1^2, x_2^2) \ldots, (x_1^{q+1}, x_2^{q+1}) \in \mathcal{X}^2$ such that $f_v(x_1^1, x_2^1) = \ldots = f_v(x_1^{q+1}, x_2^{q+1})$. Let $e = (v, u)$ be an edge adjacent to v. Here, we assume H does not have isolated vertices (otherwise they can be removed from H before the reduction, and colored by any color). By the pigeonhole principle, no other function f_u can distinguish

between $(x_1^1, x_2^1), (x_1^2, x_2^2) \ldots, (x_1^{q+1}, x_2^{q+1})$, i.e., at least two of them must have the same image under f_u, contradicting the fact that terminal $t_{(v,u)}$ can decode with zero error in 2 communication rounds. On the other hand, if there exists a function f_v and an element in $\mathcal{X}$ that has less than q sources in $\mathcal{X}^2$, then there must be another element in $\mathcal{X}$ that has at least $q+1$ sources, which implies a contradiction as presented above. Thus, all functions f_v used by the given communication scheme are balanced.

We now divide these balanced functions into $q^2 - q + 1$ classes. Assume that all elements in $\mathcal{X}^2$ are ordered, i.e., $\mathcal{X}^2 = \{(x_1^1, x_2^1), (x_1^2, x_2^2), \ldots, (x_1^{q^2}, x_2^{q^2})\}$. For $1 \leq i \leq q^2 - q + 1$, the i'th class contains all the balanced functions f for which i is the minimal index such that $f(x_1^{i+1}, x_2^{i+1}) = f(x_1^1, x_2^1)$, namely, $f(x_1^{i+1}, x_2^{i+1}) = f(x_1^1, x_2^1)$ and for all $2 \leq j \leq i$ $f(x_1^j, x_2^j) \neq f(x_1^1, x_2^1)$.

Notice that every balanced function is in exactly one of these classes. Also, to allow decoding, it holds that for any edge $e = (v, u)$, f_v, f_u cannot belong to the same class since if for some i, $f_v(x_1^i, x_2^i) = f_v(x_1^1, x_2^1)$, and $f_u(x_1^i, x_2^i) = f_u(x_1^1, x_2^1)$, terminal $t_{(v,u)}$ will not be able to distinguish between message pairs (x_1^i, x_2^i) and (x_1^1, x_2^1), as thus cannot decode with 0 error. Therefore, for all u, v for which f_v, f_u are in the same class the edge (u, v) is not in H. Thus, all vertices belonging to functions in the same class can be assigned the same color. This implies a $q^2 - q + 1$-coloring of H (assign color i to all vertices v that received a function f_v that is in the i'th class). ■

IV. Minimizing latency for given alphabet size

A. Proof of Theorem 2

We now give our proof for Theorem 2. As before, we use a reduction from rainbow coloring of hypergraphs.

Proof of Theorem 2. Let k be any constant. Given a $2k$-uniform hypergraph $H = (V_H, E_H)$ which is a hard instance of Corollary 1.1 of [38], we construct an instance $(G, S, \{X_1\}, T)$ as in the proof of Theorem 1. The vertex set of G includes a single source node r, $|V_H|$ intermediate vertices v (one vertex for every vertex in H), and $|E_H|$ vertices t_e (one vertex for every hyperedge $e \in E_H$). The edge set of G includes an edge (r, v) for all $v \in G$ corresponding to vertices in V_H and edges (v, t_e) for all $e \in E_H$, and $v \in e$. Not that here, edge e is a subset of vertices of size $2k$. The source vertex r holds k messages $x_1, x_2, \ldots x_k$ required at the terminal vertices $\{t_e\}_{e \in E_H}$. Therefore $S = \{r\}$, $X = X_1 = \{x_1, \ldots, x_k\}$, $T = \{t_e | e \in E_H\}$.

Consider a k rainbow coloring of H. We construct a 2-round communication scheme for G, over an alphabet of size q, as in the proof of Theorem 1. In the first round, r transmits on its outgoing edges $x_1, \ldots x_k$. For $1 \leq i \leq k$, for vertex v of color i in H, r will transmit x_i over edge (r, v). In the second round of communication, intermediate vertices v corresponding to vertices in V_H will forward their incoming message to the corresponding terminal vertices they are connected to. Terminal vertex t_e corresponding to edge e in H is guaranteed to receive all k messages on its incoming edges as the nodes of e are colored by k different colors in H (by its rainbow coloring).

For the other direction, assume that the instance $(G, S, \{X_1\}, T)$ is (q, τ)-feasible for $\tau \leq k$. Then there exists a τ-round communication scheme over an alphabet of size q that allows all terminals of G to decode $x_1, \ldots, x_k$. We show that H is c colorable, for $c = q^{(\tau-1)q^k}$. As in the proof of Theorem 1, we may assume without loss of generality that vertex r sends some function $f_{v,i}(x_1, \ldots, x_k)$ to intermediate vertex v in round i, $1 \leq i \leq \tau - 1$, and that every intermediate vertex v forwards $f_{v,i}(x_1, \ldots, x_k)$ to the corresponding terminals it is connected to in round $i+1$. Notice that since the total number of rounds is τ, every vertex receives $\tau - 1$ packets. We define a distinct color for each subset of size $\tau - 1$ of encoded packets. Each vertex $v \in V_H$, corresponding to an intermediate vertex in G, is colored by the color corresponding to the set of packets $\{f_{v,i}(x_1, \ldots, x_k)\}_{i=1}^{\tau-1}$. The total number of colors is at most $q^{(\tau-1)q^k}$. Since each terminal vertex t_e must decode all k packets and each node v with the same color forwards at most $\tau - 1$ encoded packets to t_e, then for $\tau \leq k$ it cannot be the case that all vertices in e are colored by the same color (e cannot be monochromatic). Thus we get a weak coloring of H using at most c colors.

Now, for constants k and $c = q^{(k-1)q^k}$, given H which is taken to be either rainbow k-colorable or alternatively not weakly c-colorable, we can distinguish between the two cases by approximating the minimum τ for which $(G, S, \{X_1\}, T)$ is (q, τ)-feasible. Namely, if we can approximate the minimum τ within a multiplicative ratio α for which $2\alpha \leq k$ we can distinguish the two cases of H using the following argument. Given H, we construct $(G, S, \{X_1\}, T)$ and approximate the minimum τ for which $(G, S, \{X_1\}, T)$ is (q, τ)-feasible. Say our approximate value is τ_0, implying that we have established that $(G, S, \{X_1\}, T)$ is (q, τ_0)-feasible. We now show that if $\tau_0 \leq k$ it must be that H was originally rainbow k-colorable, otherwise H was originally not weakly c colorable.

Specifically, if $\tau_0 \leq k$ then, using our analysis above, H is weakly $q^{(\tau_0-1)q^k}$-colorable. However, $q^{(\tau_0-1)q^k} \leq q^{(k-1)q^k} = c$, which implies that H is weakly c-colorable. Thus H must have originally been rainbow k-colorable. If $\tau_0 > k$, consider in contradiction that H was originally rainbow k-colorable. In this case, $(G, S, \{X_1\}, T)$ is $(q, 2)$-feasible, i.e., the minimum τ in this case is 2. Thus, $\tau_0 \leq 2\alpha \leq k$. A contradiction.

We conclude that approximating the minimum τ for which $(G, S, \{X_1\}, T)$ is (q, τ) feasible within an approximation ratio of $\alpha = k/2$ is NP-hard as it allows to distinguish whether H is rainbow k-colorable or alternatively not weakly c-colorable. ■

Notice in Theorem 2 above, for any $2k$-uniform hypergraph, the implied graph G is always $(2, k+1)$-feasible. Thus our inapproximability result is optimal using the reduction studied in Theorem 2.

B. Intractability of any constant approximation

We now address the question of approximating τ within any constant approximation ratio. A slight modification of the proof above shows the following corollary. Here, the term

$\omega_n(1)$ refers to any function that tends to infinity as n (the number of nodes in G) tends to infinity.

Corollary 1 *Let $(G, S, \{X_i\}_{i\in S}, T)$ be a multicast network coding instance for which $|X| = |\cup_{i\in S} X_i| = k = \omega_n(1)$. For constant q, approximating the minimum τ for which the instance is (q, τ) feasible within any constant multiplicative ratio is NP-hard.*

Proof: We use the notation from the proof of Theorem 2. Let $k = \omega_n(1)$, let ℓ be any constant greater or equal to 2, and let $H = (V_H, E_H)$ be a 2ℓ-uniform hypergraph which is a hard instance of Corollary 1.1 of [38]. Construct the graph G as in the proof of Theorem 2, and define a new graph G' that includes G, and has $k-\ell$ additional nodes $w_2, \ldots w_{k-\ell+1}$ and additional edges (w_i, t_e) for all $2 \leq i \leq k-\ell+1$, $t_e \in T$. Vertex r holds $x_1, \ldots, x_\ell$, and vertex w_i holds message $x_{\ell+i-1}$. Namely $S' = \{r, w_2, \ldots w_{k-\ell+1}\}$, $X'_1 = \{x_1, \ldots x_\ell\}$, $X'_i = \{x_{\ell+i-1}\}$, for $2 \leq i \leq k-\ell+1$. We prove that if $H = (V_H, E_H)$ has an ℓ rainbow coloring then k messages can be routed over $(G', S', \{X'_i\}_{i\in S'}, T)$ in two rounds, and if $H = (V_H, E_H)$ is not c colorable for any constant $c = q^{(\ell-1)q^\ell}$, then $(G', S', \{X'_i\}_{i\in S'}, T)$ is not (q, ℓ)-feasible:

Consider an ℓ rainbow coloring of H. We construct a 2-round communication scheme for G', over an alphabet of size q. In the first round, the source r transmits on its outgoing edges to vertices in G $x_1, \ldots x_\ell$ according the the communication scheme for G in the proof of Theorem 2, and for each $2 \leq i \leq k-\ell+1$ w_i transmits the message $x_{i+\ell-1}$ to all terminals. In the second round of communication, intermediate vertices v in G' corresponding to vertices in V_H will forward their incoming message to the corresponding terminal vertices they are connected to. Terminal vertex t_e corresponding to edge e in H is guaranteed to receive all ℓ messages on its incoming edges in G which are functions of $x_1, \ldots, x_\ell$ (as the nodes of e are colored by ℓ different colors in H) and in addition receives $x_{\ell+1}, \ldots x_k$.

For the other direction, assume that the instance $(G', S', \{X'_i\}_{i\in S'}, T)$ is (q, ℓ)-feasible. Then there exists a ℓ-round communication scheme over an alphabet of size q that allows all terminals of G' to decode $x_1, \ldots, x_k$. We show that H is c colorable, for $c = q^{(\ell-1)q^\ell}$. Note that $w_2, \ldots w_{k-\ell+1}$ must send packets that are functions of $x_{\ell+1}, \ldots, x_k$ respectively, and vertex r sends function $f_{v,i}(x_1, \ldots, x_\ell)$ to intermediate vertex v in round i, $1 \leq i \leq \ell-1$. We may assume without loss of generality as before that every intermediate vertex v forwards $f_{v,i}(x_1, \ldots, x_\ell)$ to the corresponding terminals it is connected to in round $i+1$. Following the proof of Theorem 2, we get a coloring of H using at most $q^{(\ell-1)q^\ell}$ colors. Our assertion now follows using the discussion at the end of the proof of Theorem 2 and the fact that ℓ is an arbitrarily large constant. ■

V. Concluding remarks

In this paper we studied the relation between *latency* and *alphabet size* in the context of Multicast Network Coding. For sufficiently large alphabet sizes (i.e., packet sizes), it is known that random linear multicast network coding minimizes latency. We consider the setting of non-linear codes in the finite blocklength scenario. We consider two optimization problems: minimizing alphabet size that satisfies a given latency constraint and minimizing latency that satisfies a given limit on the alphabet size. Through reductive arguments, we showed that determining or approximating the two problems is NP-hard. Roughly speaking, we note that our reductions in the proof of Theorems 1 and 2 show that given $(G, S, \{X_i\}_{i\in S}, T)$ it is NP-hard to distinguish between the case that $(G, S, \{X_i\}_{i\in S}, T)$ has a certain *high-quality* routing-based communication scheme (that does not use encoding) and the case that $(G, S, \{X_i\}_{i\in S}, T)$ does not have a *lower-quality* network-coding scheme. Here, the terms high- and low-quality are determined by the analysis appearing in Theorems 1 and 2.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.

[2] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49(2):371 – 381, 2003.

[3] R. Koetter and M. Medard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, 11(5):782 – 795, 2003.

[4] T. Ho, R. Koetter, M. Medard, D. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In *Proceedings of the IEEE International Symposium on Information Theory*, 2003.

[5] S. El Rouayheb, A. Sprintson, and P. Sadeghi. On coding for cooperative data exchange. *In Proceedings of IEEE Information Theory Workshop*, 2010.

[6] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb. A randomized algorithm and performance bounds for coded cooperative data exchange. *In Proceedings of IEEE International Symposium on Information Theory*, pages 1888–1892, 2010.

[7] D. Ozgul and A. Sprintson. An algorithm for cooperative data exchange with cost criterion. *In Proceedings of Information Theory and Applications Workshop (ITA)*, 2011.

[8] S. E. Tajbakhsh, P. Sadeghi, and R. Shams. A generalized model for cost and fairness analysis in coded cooperative data exchange. *In Proceedings of International Symposium on Network Coding (NetCod)*, 2011.

[9] T. A. Courtade and R. D. Wesel. Efficient universal recovery in broadcast networks. *In Proceedings of Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing*, 2010.

[10] T.A. Courtade, B. Xie, and R. Wesel. Optimal exchange of packets for universal recovery in broadcast networks. *In Proceedings of Military Communications Conference*, 2010.

[11] N. Aboutorab, S. Sadeghi, and S. E. Tajbakhsh. Instantly decodable network coding for delay reduction in cooperative data exchange systems. In *Proceedings of IEEE International Symposium on Information Theory*, pages 3095 – 3099, 2013.

[12] S. E. Tajbakhsh, S. Sadeghi, and N. Aboutorab. Instantly decodable network codes for cooperative index coding problem over general topologies. In *Proceedings of Communications Theory Workshop (AusCTW)*, pages 84 – 89, 2014.

[13] T.A. Courtade and R.D. Wesel. Coded cooperative data exchange in multihop networks. *IEEE Transactions on Information Theory*, 60(2):1136 – 1158, 2014.

[14] M. Gonen and M. Langberg. Coded cooperative data exchange problem for general topologies. *IEEE Transactions on Information Theory*, 61(10):5656–5669, 2015.

[15] S. Deb, M. Médard, and C. Choute. Algebraic gossip: A network coding approach to optimal multiple rumor mongering. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2486–2507, 2006.

[16] D. Mosk-Aoyama and D. Shah. Information dissemination via network coding. In *IEEE International Symposium on Information Theory*, pages 1748–1752, 2006.

[17] M. Borokhovich, C. Avin, and Z. Lotker. Tight bounds for algebraic gossip on graphs. In *IEEE International Symposium on Information Theory*, pages 1758–1762, 2010.

[18] C. Avin, M. Borokhovich, K. Censor-Hillel, and Z. Lotker. Order optimal information spreading using algebraic gossip. *Distributed computing*, 26(2):99–117, 2013.

[19] A. Cohen, B. Haeupler, C. Avin, and M. Médard. Network coding based information spreading in dynamic networks with correlated data. *IEEE Journal on Selected Areas in Communications*, 33(2):213–224, 2015.

[20] B. Tang, B.u Ye, S.g Guo, S. Lu, and D. O. Wu. Order-optimal information dissemination in manets via network coding. *IEEE Transactions on Parallel and Distributed Systems*, 25(7):1841–1851, 2014.

[21] B. Haeupler. Analyzing network coding (gossip) made easy. *Journal of the ACM (JACM)*, 63(3):26, 2016.

[22] B. Haeupler and D. Karger. Faster information dissemination in dynamic networks via network coding. In *Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed computing*, pages 381–390, 2011.

[23] B. Haeupler and F. Kuhn. Lower bounds on information dissemination in dynamic networks. In *International Symposium on Distributed Computing*, pages 166–180. Springer, 2012.

[24] C. C. Wang and M. Chen. Sending perishable information: Coding improves delay-constrained throughput even for single unicast. In *IEEE International Symposium on Information Theory*, pages 866–870, 2014.

[25] C. Chekuri, S. Kamath, S. Kannan, and P.d Viswanath. Delay-constrained unicast and the triangle-cast problem. In *IEEE International Symposium on Information Theory (ISIT)*, pages 804–808, 2015.

[26] B. Haeupler, M. Kim, and M. Médard. Optimality of network coding with buffers. In *IEEE Information Theory Workshop (ITW)*, pages 533–537, 2011.

[27] April Rasala Lehman and Eric Lehman. Complexity classification of network information flow problems. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 142–150. Society for Industrial and Applied Mathematics, 2004.

[28] Søren Riis. Linear versus non-linear boolean functions in network flow. In *38th Annual Conference on Information Science and Systems (CISS), Princeton, NJ*, 2004.

[29] Randall Dougherty, Christopher Freiling, and Kenneth Zeger. Linearity and solvability in multicast networks. *IEEE Transactions on Information Theory*, 50(10):2243–2256, 2004.

[30] Qifu Tyler Sun, Xunrui Yin, Zongpeng Li, and Keping Long. Multicast network coding and field sizes. *IEEE Transactions on Information Theory*, 61(11):6182–6191, 2015.

[31] C. Fragouli and E. Soljanin. Network Coding Fundamentals. *Foundations and Trends in Networking*, 2(1):1–133, 2007.

[32] Michael Langberg. Graph coloring. In *Encyclopedia of Algorithms*, pages 368–371. Springer, 2008.

[33] U. Feige and J. Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57:187–199, 1998.

[34] J. Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Math*, 182(1):105–142, 1999.

[35] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the 38th annual ACM symposium on Theory of Computing*, pages 681–690, 2006.

[36] V. Guruswami and S. Khanna. On the Hardness of 4-coloring a 3-colorable Graph. *Proceedings of the 15th Annual IEEE Conference on Computational Complexity*, pages 188–197, 2000.

[37] S. Khanna, N. Linial, and S. Safra. On the Hardness of Approximating the Chromatic Number. *Combinatorica*, 20:393–415, 2000.

[38] V. Guruswami and E. Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 822–836, 2015.

[39] I. Dinur, O. Regev, and C.D. Smyth. The hardness of 3 - uniform hypergraph coloring. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, 2002.

[40] L. Lovàsz. Colorings and coverings of hypergraphs. In *Proceedings of the 4th Southeastern Conference on Combinatorics, Graph Theory, and Computing, Utilitas Mathematica Publishing, Winnipeg*, pages 3–12, 1973.