

2018

Zero-error Function Computation on a Directed Acyclic Network

Ardhendu Tripathy
Iowa State University

Aditya Ramamoorthy
Iowa State University, adityar@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/ece_pubs



Part of the [Systems and Communications Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/ece_pubs/181. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Electrical and Computer Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Zero-error Function Computation on a Directed Acyclic Network

Abstract

We study the rate region of variable-length source-network codes that are used to compute a function of messages observed over a network. The particular network considered here is the simplest instance of a directed acyclic graph (DAG) that is not a tree. Existing work on zero-error function computation in DAG networks provides bounds on the $\text{computation capacity}$, which is a measure of the amount of communication required per edge in the worst case. This work focuses on the average case: an achievable rate tuple describes the expected amount of communication required on each edge, where the expectation is over the probability mass function of the source messages. We describe a systematic procedure to obtain outer bounds to the rate region for computing an arbitrary demand function at the terminal. Our bounding technique works by lower bounding the entropy of the descriptions observed by the terminal conditioned on the function value and by utilizing the Schur-concave property of the entropy function. We evaluate these bounds for certain example demand functions.

Disciplines

Electrical and Computer Engineering | Systems and Communications

Comments

This is a manuscript of Tripathy, Ardhendu, and Aditya Ramamoorthy. "Zero-error Function Computation on a Directed Acyclic Network." *arXiv preprint arXiv:1805.03730* (2018).

Zero-error Function Computation on a Directed Acyclic Network

Ardhendu Tripathy, *Student Member, IEEE*, and Aditya Ramamoorthy, *Member, IEEE*

Abstract

We study the rate region of variable-length source-network codes that are used to compute a function of messages observed over a network. The particular network considered here is the simplest instance of a directed acyclic graph (DAG) that is not a tree. Existing work on zero-error function computation in DAG networks provides bounds on the *computation capacity*, which is a measure of the amount of communication required per edge in the worst case. This work focuses on the average case: an achievable rate tuple describes the expected amount of communication required on each edge, where the expectation is over the probability mass function of the source messages.

We describe a systematic procedure to obtain outer bounds to the rate region for computing an arbitrary demand function at the terminal. Our bounding technique works by lower bounding the entropy of the descriptions observed by the terminal conditioned on the function value and by utilizing the Schur-concave property of the entropy function. We evaluate these bounds for certain example demand functions.

I. INTRODUCTION

Computing functions of data observed over a network is a well-motivated problem, and different frameworks addressing the problem have been studied in the literature. Broadly speaking, a general function computation problem can be modeled in the following manner. A directed acyclic graph (DAG) is used to model a communication network. Its vertices denote the nodes of the network that are assumed to have unrestricted computational power and storage capacity. The edges denote one-way communication links that can be thought of as noiseless *bit-pipes*. Each node can act as a decoder on the information received on its incoming edges and as an encoder for information transmitted on its outgoing edges. Some of the nodes in the network, called the *source* nodes, observe discrete-valued source *messages* that take values in a finite alphabet. The random process generating the source messages is assumed to be stationary and memoryless. There is a single *terminal* node that wishes to compute losslessly a discrete-valued function of all the source messages using the information it receives on its incoming edges. A solution to such a function computation problem will specify the communication carried out on each link and the information processing performed at each node of the network. We are interested in finding the minimal communication required for solving a given problem instance. The amount of communication in a network can be specified by a *rate tuple* that has as many entries as the number of edges in the network. Each entry denotes the rate of the code employed on the corresponding edge in the network. As remarked in [1], this problem in its full generality encompasses several different areas in information theory, and as such, existing literature focuses on simplified versions that highlight different aspects of the problem.

If there is just one encoder and one decoder connected by a noiseless communication link, and the decoder wants to compute the identity function for the message, then it is a standard source coding problem. When there is some coded side information available through an additional link at the decoder, then the optimal rate pair is given by the Ahlswede–Körner–Wyner solution [2, Thm. 10.2]. Extending this one-help-one solution to a two-help-one scenario is not optimal, as was demonstrated for a particular two-help-one problem instance by Körner and Marton [3].

Consider now the case of two encoders, connected by two separate links to a decoder which is interested in computing the identity function on the pair of source messages. The optimal rate pair for this distributed source coding problem is known to be the Slepian–Wolf rate region, and this solution can be extended to multiple encoders each of which are directly connected to the decoder via a separate link [2, Chap. 10].

In reproducing the source messages at the terminal in the above scenarios, the coding schemes allow for an ϵ -block error that can be made as small as desired by choosing asymptotically large block lengths. Concurrently, work on zero-error source coding with side information was initiated by Witsenhausen [4]. With X being the source message and Y the side information, he defined a *confusability graph* G_X for X that specifies the realizations of X that must be given distinct codewords by the encoder in order to attain zero-error. The optimum number of distinct codewords required was shown to be the graph chromatic number $\chi(G_X)$; for encoding multiple (say, k) instances it is $\chi(G_X^k)$ where G_X^k denotes the k -wise AND product graph of G_X . Thus when $\chi(G_X^k) < (\chi(G_X))^k$, block encoding multiple instances allows one to reduce the number of distinct codewords required. The above idea along with the p.m.f. of X was used to find the expected number of bits that must be transmitted by the encoder in [5]. For the case when the codewords are restricted to be prefix-free, their expected length was shown in [5] to be within one bit of the chromatic entropy $H_\chi(G_X, X)$ of the probabilistic graph (G_X, X) . When multiple

This work was supported in part by the National Science Foundation (NSF) under grants CCF-1149860, CCF-1320416, CCF-1718470 and DMS-1120597. The material in this work has appeared in part at the 2016 IEEE International Symposium on Information Theory. The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50011 USA. e-mails: {ardhendu, adityar}@iastate.edu

instances are block encoded, the asymptotic per-instance expected codeword length is $\lim_{k \rightarrow \infty} \frac{1}{k} H_{\mathcal{X}}(G_X^{\vee k}, X^k)$, where $G_X^{\vee k}$ is the k -wise OR product graph of G_X . This limit was shown to be the *graph entropy* of (G_X, X) in [5]. More information about the possible savings due to encoding multiple instances can be found in [6, Sec. XI].

In a function computing problem, the terminal would want to compute a general demand function of the messages and not necessarily the identity function. Reference [7] considered computing a function on the setup of the source coding with side information problem, and allowed a vanishing block-error probability. They defined the conditional graph entropy $H(G_X, X|Y)$ for the probabilistic graph $(G_X, X|Y)$ of X given Y and showed that it is equal to the optimal number of bits per-instance of X that must be communicated by the encoder when asymptotically large block lengths are used. This was extended by the authors in [8] where they defined a conditional chromatic entropy of a probabilistic graph and showed that $\lim_{k \rightarrow \infty} \frac{1}{k} H_{\mathcal{X}}(G_X^{\vee k}, X^k|Y^k) = H(G_X, X|Y)$. This gave a graph coloring procedure to obtain codes with rate close to the lower bound of [7].

Function computation was also considered in the distributed setting where two encoders separately encode X and Y such that the decoder is able to compute $F(X, Y)$ losslessly. This scenario is closer to the source coding with side information problem than the distributed source coding problem because of the following. Any demand function can be computed at the decoder after communicating each of the source messages to it, thus the Slepian–Wolf region is an inner bound to the rate region of a function computation problem on the same network. Körner and Marton showed in [3] that if the decoder wants to decode just Z , then the rate tuple $(R_X, R_Y, R_Z) = (H(Z), H(Z), 0)$ is achievable for the case when $Z = X \oplus Y$, where X and Y are a doubly symmetric binary source pair and \oplus denotes modulo-2 sum. This rate tuple, with $R_Z = 0$, can be interpreted as the decoder wanting to compute the modulo-2 sum using the information obtained from the X and Y encoders. This view was taken by Han and Kobayashi in [9] where, subject to the constraint $\Pr(X = x, Y = y) > 0, \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, they gave necessary and sufficient conditions on the demand function $F(X, Y)$ for which the Slepian–Wolf region for the source pair (X, Y) coincides with the rate region of the function computation problem.

The authors in [8] also considered distributed function computation, where they showed that if the joint p.m.f $\Pr(X = x, Y = y)$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ satisfied a restrictive ‘zigzag’ condition, then coloring the confusability graphs $G_X^{\vee k}$ and $G_Y^{\vee k}$ and then using Slepian–Wolf code is optimal for asymptotic block length. This sufficient condition was relaxed in [10] to a coloring connectivity condition by taking into account the function value. This condition was shown to characterize the rate region for function computation on one-stage tree-networks. It also gave an inner bound to the rate region of a general tree network. The rate region for a multi-stage tree-network when every source message at a vertex in the network satisfies a local Markovian property was characterized in [11]. The function computation scenarios described above all allowed for a vanishing block-error probability with asymptotically large block lengths.

In this paper, we study the problem of zero-error function computation over a simple DAG network that is not a tree, shown in Figure 1. We assume that the three source messages are independent and the terminal wants to compute an arbitrary specified demand function of the messages. We allow for block encoding and decoding of multiple instances at the sources and the terminal. Even with the independence assumption, characterizing the rate region of this function computation problem is difficult as was observed in [21], where the problem of computing a particular *arithmetic sum* demand function on the same DAG network was considered. We describe works closely related to our problem next.

A. Related work

Zero-error function computation over a graphical network using network coding [12], [13] was studied in [14]. There they considered two variants of the communication load on the network, called *worst-case* and *average-case* complexity, depending on whether the probability information of the source messages was used or not. They characterized the rate region of achievable rate tuples that allowed zero-error function computation in tree-networks, each entry in a rate tuple was the rate of a code employed on the corresponding edge in the tree-network. They also made the observation that finding the rate region of a DAG network is significantly challenging because of multiple paths between a source node and the terminal, which allows for different ways of combining information at the intermediate nodes.

Zero-error function computation was also studied in [15], where the authors defined the *computation capacity* of a function computation problem instance. This is a generalization of the coding capacity of a network (*c.f.* [16, Sec. VI], [17]), which is the supremum of the ratio $\frac{k}{n}$ over all achievable (k, n) fractional coding solutions for that communication network. A (k, n) fractional network code is one in which k source messages are block encoded at each encoder and every edge in the network transmits n symbols from the alphabet in one channel use. The authors in [15] characterized the computation capacity of multi-stage tree-networks by finding the necessary and sufficient amount of information that must be transmitted across all graph cuts that separate one or more source nodes from the terminal. Upper bounds on the computation capacity of DAG networks are more complicated and have been obtained in [18], [19]. These upper bounds were shown to be unachievable for a function computation problem on a particular DAG in recent work [25, Sec. V]. An illuminating example in the search for improved upper bounds for function computation on DAG networks has been the problem of computing the arithmetic sum of three source bits over the network shown in figure 1. By counting the necessary and sufficient number of codewords required, the computation capacity for this problem was evaluated to be $\log_6 4$ in [15, Sec. V]. A qualitatively different line of work considers the computation of simple functions such as finite-field sum [27]–[29] over arbitrary acyclic networks with multiple

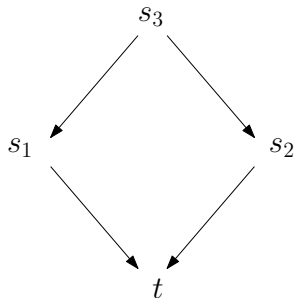


Fig. 1. A directed acyclic network with three sources, two of which also act as relay nodes, and one terminal.

terminals, i.e., there is a set of terminals each of which is interested in recovering a function of certain source messages (see also [30]–[32]) that discuss the multiple unicast problem which is an important special case of function computation). The literature in this area examines the role of the network topology in function computation.

While the upper bounds for DAG networks described above hold for the worst-case scenario, the arithmetic sum example (cf. Example 1 in Section II) shows that we can do better in the average-case scenario by using the probability information of the source messages. It can be seen that an upper bound to the computation capacity corresponds to an outer bound to the rate region of a function computation problem. We use the framework of a *source-network* code as used for zero-error network coding in [22]. Correspondingly, the quantity of interest here is the zero-error function computation rate region, and we provide outer bounds to this rate region for computing an arbitrary specified demand function on the network shown in figure 1. We summarize our contributions below.

B. Main contributions

- We compute lower bounds on the rates that must be used on the edges of the DAG in Figure 1 in order to compute with zero-error an arbitrary demand function of the messages at the terminal. This provides us an outer bound to the achievable rate region.
- The technique used for obtaining the lower bounds involves lower bounding the conditional entropy of the descriptions transmitted on the edges given the demand function value. This is done by first finding a family of p.m.f.s. that must necessarily contain the conditional p.m.f. of the descriptions transmitted by any valid source-network code for the problem. The required lower bound was then obtained by finding the entropy-minimizing p.m.f. in this family. This p.m.f. was found using the Schur-concave property of the entropy function.
- Computing the arithmetic sum over the DAG network in figure 1 was considered in [21]. By letting the number of symbols transmitted across the edges be a random variable N (instead of a fixed n for every k source bits as in the (k, n) fractional network code framework¹), it was shown that the computation rate $\frac{k}{\mathbb{E}N} = 0.8$ is achievable. This is larger than its computation capacity in the (k, n) fractional code framework, which was evaluated to be $\log_6 4$ in [15]. The work in [21] also gave an upper bound of $8/9$ for the computation capacity of this arithmetic sum example in the framework of *variable-length* network codes. We recover and improve upon their results in the rate region framework, by obtaining a tighter outer bound to rate region in this case.
- When the demand function to be computed over the DAG in figure 1 is set to be the $GF(2)$ -sum, we show using a simple achievable scheme that the outer bound obtained for the rate region is tight.

The paper is organized as follows. Section II describes the problem setup formally and motivates the use of variable-length network codes as defined in Section II-A. Section III describes the procedure used to obtain outer bounds to the rate region for computing an arbitrary demand function over the network in Figure 1. Central to our approach are lower bounds to the conditional entropy of the descriptions transmitted, and these are described and illustrated using a running example in Section III-A. We also consider two other example demand functions in Sections III-D and III-C. Section IV concludes the chapter and lists some avenues for future work.

II. PROBLEM FORMULATION

The edges in Figure 1 (later denoted by an ordered pair of vertices) have unit-capacity, i.e., they transmit one symbol from \mathcal{Z} in one time slot. We use the notation of a standard network code from [15]. In what follows, all logarithms denoted as \log are to the base 2 unless specified otherwise. Suppose that \mathcal{Z} is the alphabet used for communication, and $|\mathcal{Z}| > 1$. Vertices s_1, s_2, s_3 are the three memoryless source nodes that observe source messages X_1, X_2, X_3 respectively, each from a discrete

¹As described in [21], for finding the computation capacity, it can be assumed w.l.o.g. that the message vector X_3^k is available to both the encoders at s_1 and s_2 . Then the random variable N is defined as a stopping time w.r.t. the pair of random variables transmitted over the edges (s_1, t) and (s_2, t) .

alphabet \mathcal{A} with size $|\mathcal{A}| > 1$. The sources are assumed to be i.i.d. uniformly distributed over \mathcal{A} . Terminal node t wants to compute with zero error and zero distortion a known demand function:

$$f : \mathcal{A} \times \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{B},$$

where \mathcal{B} is a discrete alphabet of size $|\mathcal{B}| > 1$. To avoid trivialities we assume that the demand function is not constant in any of its three arguments. A (k, n) network code that satisfies the terminal has the following components.

- An *encoding function* $h^e(\cdot)$ associated with every edge e in the network:

$$\begin{aligned} h^{(s_3, s_1)}(X_3^k) &: \mathcal{A}^k \rightarrow \mathcal{Z}^n \\ h^{(s_3, s_2)}(X_3^k) &: \mathcal{A}^k \rightarrow \mathcal{Z}^n \\ h^{(s_1, t)}(X_1^k, h^{(s_3, s_1)}(X_3^k)) &: \mathcal{A}^k \times \mathcal{Z}^n \rightarrow \mathcal{Z}^n \\ h^{(s_2, t)}(X_2^k, h^{(s_3, s_2)}(X_3^k)) &: \mathcal{A}^k \times \mathcal{Z}^n \rightarrow \mathcal{Z}^n \end{aligned}$$

Here, the notation $X_j^k, j \in \{1, 2, 3\}$ denotes a block of k i.i.d. uniform *messages* observed at source node s_j .

- A *decoding function* $\psi(\cdot)$ used by the terminal t :

$$\psi\left(h^{(s_1, t)}(X_1^k, h^{(s_3, s_1)}(X_3^k)), h^{(s_2, t)}(X_2^k, h^{(s_3, s_2)}(X_3^k))\right) : \mathcal{Z}^n \times \mathcal{Z}^n \rightarrow \mathcal{B}^k$$

With slight abuse of notation we let $f(X_1^k, X_2^k, X_3^k) \in \mathcal{B}^k$ denote the k values returned by the demand function $f(X_1, X_2, X_3)$ when applied component-wise on a block of k i.i.d. (X_1, X_2, X_3) triples. That is,

$$f(X_1^k, X_2^k, X_3^k) = \left(f(X_1^{(1)}, X_2^{(1)}, X_3^{(1)}), f(X_1^{(2)}, X_2^{(2)}, X_3^{(2)}), \dots, f(X_1^{(k)}, X_2^{(k)}, X_3^{(k)})\right),$$

where $X_j^{(i)}$ denotes the i th component of X_j^k . By the zero-error criterion we have that

$$\Pr\left\{\psi\left(h^{(s_1, t)}(X_1^k, X_3^k), h^{(s_2, t)}(X_2^k, X_3^k)\right) \neq f(X_1^k, X_2^k, X_3^k)\right\} = 0.$$

If such a set of encoding and decoding functions exist for a choice of positive integers k and n , we say that the network code *computes* the demand function and has *computation rate* $= k \log |\mathcal{A}| / (n \log |\mathcal{Z}|)$. The *computation capacity* for a particular demand function is defined to be the supremum of all achievable computation rates.

The above framework is restrictive in the sense that the block length of the network code used (i.e., the value of n) is the same for each edge in the network. If we know the probability distribution associated with the message random variables and allow the block length of the network code on edge e to be a random variable N_e , then we can compress the descriptions transmitted on the edges and obtain savings in the expected length $\mathbb{E}[N_e]$.

Example 1: Consider the problem of computing the sum over the integers (arithmetic sum) of three messages, i.e. $f(X_1, X_2, X_3) = X_1 + X_2 + X_3$, on the network shown in Figure 1. The messages $X_1, X_2, X_3 \in \{0, 1\}^k$ and $f(X_1, X_2, X_3) \in \{0, 1, 2, 3\}^k$. This example was first considered in [15], where they solve it using an optimal network code with $\mathcal{Z} = \{0, 1\}$ that has computation rate $\log_6 4 \approx 0.77$. The optimal network code transmits the value of X_3 on the edges (s_3, s_1) , (s_3, s_2) and the values transmitted on the other two edges are as follows, assuming block length k to be a multiple of 2.

$$\begin{aligned} h^{(s_1, t)}(X_1, X_3) &= \begin{cases} X_1^{(i)} + X_3^{(i)}, & \text{for all } 1 \leq i \leq k/2 \\ X_1^{(i)}, & \text{otherwise.} \end{cases} \\ h^{(s_2, t)}(X_2, X_3) &= \begin{cases} X_2^{(i)}, & \text{for all } 1 \leq i \leq k/2 \\ X_2^{(i)} + X_3^{(i)}, & \text{otherwise.} \end{cases} \end{aligned}$$

The number of bits to be transmitted on the edges using the above encoding functions can be seen to be $k(1 + \log 3)/2$. Now suppose that the messages are such that each $X_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(0.5)$, i.e., they are equally likely random bits. Then both $X_1^{(i)} + X_3^{(i)}$ and $X_2^{(i)} + X_3^{(i)}$ in the encoding functions above have a biased distribution and can be compressed. The entropy $H(X_1^{(i)} + X_3^{(i)}) = 1.5$, and hence its ϵ -typical set [23] for $k/2$ instances can be enumerated using $1.5k/2$ bits. Thus the expected number of bits needed to be transmitted in this case can be seen to be

$$\mathbb{E} N_{(s_1, t)} = \mathbb{E} N_{(s_2, t)} = (1 - \epsilon)(3k/4 + k/2) + \epsilon(k \log_2 3 + k/2) \approx 5k/4,$$

where the ϵ can be made small enough using large k . We note that $5/4 < (1 + \log 3)/2$, and that an analogous definition for the computation rate of a network code in the variable-length framework would give that $k / \max\{\mathbb{E}[N_{(s_1, t)}], \mathbb{E}[N_{(s_2, t)}]\} = 0.8 > 0.77$.

Example 2: Consider computing the maximum over the integers $f(X_1, X_2) = \max\{X_1, X_2\}$, where $X_1, X_2 \in \{0, 1\}$, over the reverse butterfly network shown in Figure 2. This problem was considered in [25], and they gave an upper bound of 2 for

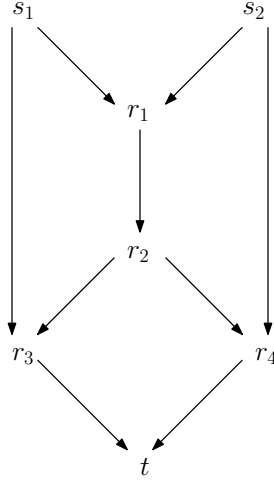


Fig. 2. A directed acyclic network with two sources, four relay nodes and one terminal.

the computation rate of any valid network code. They also gave a valid network code with rate $3/2$ that has the following edge functions, assuming block length k to be a multiple of 3.

$$\begin{aligned}
h^{(s_1, r_3)}(X_1^k) &= (X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(k/3)}), \\
h^{(s_1, r_1)}(X_1^k) &= (X_1^{(1+k/3)}, X_1^{(2+k/3)}, \dots, X_1^{(k)}), \\
h^{(s_2, r_4)}(X_2^k) &= (X_2^{(1)}, X_2^{(2)}, \dots, X_2^{(k/3)}), \\
h^{(s_2, r_1)}(X_2^k) &= (X_2^{(1+k/3)}, X_2^{(2+k/3)}, \dots, X_2^{(k)}), \\
h^{(r_1, r_2)}(h^{(s_1, r_1)}(X_1^k), h^{(s_2, r_1)}(X_2^k)) &= (\max\{X_1^{(1+k/3)}, X_2^{(1+k/3)}\}, \dots, \max\{X_1^{(k)}, X_2^{(k)}\}), \\
h^{(r_2, r_3)}(h^{(r_1, r_2)}(h^{(s_1, r_1)}(X_1^k), h^{(s_2, r_1)}(X_2^k))) &= (\max\{X_1^{(1+k/3)}, X_2^{(1+k/3)}\}, \dots, \max\{X_1^{(2k/3)}, X_2^{(2k/3)}\}), \\
h^{(r_2, r_4)}(h^{(r_1, r_2)}(h^{(s_1, r_1)}(X_1^k), h^{(s_2, r_1)}(X_2^k))) &= (\max\{X_1^{(1+2k/3)}, X_2^{(1+2k/3)}\}, \dots, \max\{X_1^{(k)}, X_2^{(k)}\}), \\
h^{(r_3, t)}(h^{(s_1, r_3)}(\cdot), h^{(r_2, r_3)}(\cdot)) &= (h^{(s_1, r_3)}(\cdot), h^{(r_2, r_3)}(\cdot)), \\
h^{(r_4, t)}(h^{(s_2, r_4)}(\cdot), h^{(r_2, r_4)}(\cdot)) &= (h^{(s_2, r_4)}(\cdot), h^{(r_2, r_4)}(\cdot)),
\end{aligned}$$

where the last two equations state that the edges (r_3, t) , (r_4, t) concatenate their inputs and forward it, and we have omitted the arguments of certain edge functions (represented by a dot) for brevity.

Now consider the case when each $X_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(0.75)$, i.e., it takes the value 1 with probability 0.75 and 0 with probability 0.25. Since $\max\{X_1^{(i)}, X_2^{(i)}\}$ has a biased p.m.f., we are able to compress the descriptions transmitted on the edges. We describe a variable-length network code that has a similar structure, except that the partitioning of the k components of X_1 along the edges (s_1, r_3) and (s_1, r_1) is different from the 1:2 ratio used before, and it uses typical set encoding. For the i th component, the entropy values $H(X_j^{(i)}) \approx 0.8113$ and $H(\max\{X_1^{(i)}, X_2^{(i)}\}) \approx 0.3373$ can be verified. Set the value $c \triangleq 1/(2 - 0.3373/(2 \cdot 0.8113)) \approx 0.558$. Then the description transmitted on the edges (s_1, r_3) and (s_1, r_1) are as follows.

$$\begin{aligned}
h^{(s_1, r_3)}(X_1) &= \mathcal{T}_\epsilon[(X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(k-ck)})] \\
h^{(s_1, r_1)}(X_1) &= \mathcal{T}_\epsilon[(X_1^{(1+ck)}, X_1^{(2+ck)}, \dots, X_1^{(k)})],
\end{aligned}$$

where $\mathcal{T}_\epsilon[\cdot]$ indicates typical set encoding and a similar description of X_2 is transmitted on the edges (s_2, r_1) and (s_2, r_4) . This partitioning is chosen so as to have the best possible rate within a class of network codes that have the same structure as the initial network code described, as the following computations indicate.

$$\begin{aligned}
\mathbb{E} N_{(s_1, r_1)} &= \mathbb{E} N_{(s_2, r_1)} = 0.8113 \cdot ck = 0.4527k, \quad \mathbb{E} N_{(s_1, r_3)} = \mathbb{E} N_{(s_2, r_4)} = 0.8113 \cdot (1-c)k = 0.3586k, \\
\mathbb{E} N_{(r_1, r_2)} &= 0.3373 \cdot ck = 0.1882k, \quad \mathbb{E} N_{(r_2, r_3)} = \mathbb{E} N_{(r_2, r_4)} = 0.1882k/2 = 0.0941k, \\
\mathbb{E} N_{(r_3, t)} &= \mathbb{E} N_{(r_4, t)} = 0.0941k + 0.3586k = 0.4527k.
\end{aligned}$$

Thus we have that $k/(\max_{\text{all edges } e} \mathbb{E}[N_e]) = k/(0.4527k) = 2.209 > 2$.

The above examples illustrate the reduction in communication load that can be achieved by using the knowledge of the probability distribution of the messages. The objective in this paper is to find bounds on rates of variable-length network codes

that are valid for a given function computation problem. We first define the framework of variable-length network codes as adapted to the network in Figure 1 next.

A. Variable-length network code for network in Figure 1

We use the *Source-Network Code* framework as described in [22] and adapt it to the function computation setting described above. The quantity of interest here is the rate region \mathcal{R} , which is a region containing all achievable rate tuples \mathbf{R} . Each rate tuple has four components, one for each edge in the network. We define the source-network code and the admissible rate tuples below.

Definition 1: Let \mathcal{Z}^* denote the set of all finite-length sequences with alphabet \mathcal{Z} . A source-network code $\mathcal{C}_{f,k}$ for computing $f(X_1^k, X_2^k, X_3^k)$ in the network of Figure 1 has the following components:

- 1) Encoding functions for edges $e \in \{(s_3, s_1), (s_3, s_2), (s_1, t), (s_2, t)\}$:

$$\begin{aligned}\phi_{(s_3, s_1)}(X_3^k) &: \mathcal{A}^k \rightarrow \mathcal{Z}^* \\ \phi_{(s_3, s_2)}(X_3^k) &: \mathcal{A}^k \rightarrow \mathcal{Z}^* \\ \phi_{(s_1, t)}(X_1^k, \phi_{(s_3, s_1)}(X_3^k)) &: \mathcal{A}^k \times \mathcal{Z}^* \rightarrow \mathcal{Z}^* \\ \phi_{(s_2, t)}(X_2^k, \phi_{(s_3, s_2)}(X_3^k)) &: \mathcal{A}^k \times \mathcal{Z}^* \rightarrow \mathcal{Z}^*\end{aligned}$$

For brevity, we denote $\phi_{(s_1, t)}(X_1^k, \phi_{(s_3, s_1)}(X_3^k))$ by the random variable \mathbf{Z}_1 , and similarly define the r.v.s $\mathbf{Z}_2, \mathbf{Z}_{31}, \mathbf{Z}_{32}$.

- 2) Decoding function for terminal t : $\psi_t: \mathcal{Z}^* \times \mathcal{Z}^* \rightarrow \mathcal{B}^k$ is such that $\Pr\{\psi_t(\mathbf{Z}_1, \mathbf{Z}_2) \neq f(X_1^k, X_2^k, X_3^k)\} = 0$.

Thus the outputs of the encoders are variable length, and the terminal is equipped with a decoder that takes in a pair of variable length inputs and returns without any error the block of k function computations on the message tuple. The rate of a source-network code is defined below, taking into account the different alphabets in which the messages and the codewords reside.

Definition 2: $\mathbf{R} = (R_{31}, R_{32}, R_1, R_2)$ is an admissible rate pair for the code $\mathcal{C}_{f,k}$ if for any $\epsilon > 0$ there exists a sufficiently large k such that

$$\log |\mathcal{Z}| \mathbb{E} \ell(\mathbf{Z}_1) \leq k \log |\mathcal{A}| (R_1 + \epsilon)$$

where $\mathbb{E} \ell(\mathbf{Z}_1)$ is the expected length (in symbols from \mathcal{Z} , and over the probability space of all message realizations) of the codeword \mathbf{Z}_1 . A similar definition is used for the rates R_{31}, R_{32} and R_2 .

III. BOUNDS ON THE RATE REGION FOR NETWORK IN FIGURE 1

We use a lower bound on the expected length of the codewords transmitted on the edges in terms of their entropy. Theorem 3 in [24] gives a lower bound for the expected length of a non-singular code over binary alphabet. The proof of the following lemma, in Appendix A, adapts their proof procedure for codes over non-binary alphabet \mathcal{Z} .

Lemma 1 (Adapted from Theorem 3 in [24]): The expected length (in symbols from an alphabet \mathcal{Z}) of the best non-singular code $\mathcal{C}_{\text{NS}}^*(\mathbf{Z})$ for a random variable \mathbf{Z} satisfies the following lower bound:

$$\mathbb{E} \ell(\mathcal{C}_{\text{NS}}^*(\mathbf{Z})) \geq H_{|\mathcal{Z}|}(\mathbf{Z}) - 2 \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}) + |\mathcal{Z}|).$$

Since the identity mapping is also a non-singular code for \mathbf{Z} , we have that

$$\mathbb{E} \ell(\mathbf{Z}) = \sum_{\mathbf{z}} \Pr\{\mathbf{Z} = \mathbf{z}\} \ell(\mathbf{z}) \geq \mathbb{E} \ell(\mathcal{C}_{\text{NS}}^*(\mathbf{Z})). \quad (1)$$

Because of the zero-error requirement for the decoding function at the terminal, we can give a value for what the sum rate $R_{31} + R_{32}$ must be greater than.

Lemma 2: Consider an equivalence relation on \mathcal{A}^k for which $\mathbf{x}_3 \equiv \mathbf{x}'_3$ if and only if for all $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{A}^k \times \mathcal{A}^k$, we have that $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}'_3)$. Define the function $g(X_3^k)$ which returns the equivalence class that X_3^k belongs to under the above relation. Then the range of $g(X_3^k)$ is a subset of $\{1, 2, \dots, |\mathcal{A}|^k\}$ and we have that $R_{31} + R_{32} \geq H(g(X_3^k))/k \log |\mathcal{A}|$.

Proof: Suppose that $H_{|\mathcal{Z}|}(g(X_3^k)|\mathbf{Z}_{31}, \mathbf{Z}_{32}) > 0$, then one cannot obtain $g(X_3^k)$ from the pair $(\mathbf{Z}_{31}, \mathbf{Z}_{32})$, i.e., there exist $\mathbf{x}_3 \not\equiv \mathbf{x}'_3$ but their associated codewords satisfy $\mathbf{z}_{31} = \mathbf{z}'_{31}$ and $\mathbf{z}_{32} = \mathbf{z}'_{32}$. There exists a pair $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{A}^k \times \mathcal{A}^k$ such that $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \neq f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}'_3)$. However, since $\mathbf{z}_{31} = \mathbf{z}'_{31}$ and $\mathbf{z}_{32} = \mathbf{z}'_{32}$, the codewords transmitted on the edges $(s_1, t), (s_2, t)$ in the two cases satisfy $\mathbf{z}_1 = \mathbf{z}'_1$ and $\mathbf{z}_2 = \mathbf{z}'_2$. Thus the decoder receives the same input arguments in both the cases and consequently causes an error.

Thus we have that $H_{|\mathcal{Z}|}(g(X_3^k)|\mathbf{Z}_{31}, \mathbf{Z}_{32}) = 0$. That gives us $H_{|\mathcal{Z}|}(g(X_3^k)) \leq H_{|\mathcal{Z}|}(\mathbf{Z}_{31}) + H_{|\mathcal{Z}|}(\mathbf{Z}_{32})$, and using the upper bound to the entropy in terms of the expected codeword length (c.f. equation (1) and lemma 1), we have the following.

$$\begin{aligned}H_{|\mathcal{Z}|}(g(X_3^k)) &\leq \mathbb{E} \ell(\mathbf{Z}_{31}) + 2 \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_{31}) + |\mathcal{Z}|) + \mathbb{E} \ell(\mathbf{Z}_{32}) + 2 \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_{32}) + |\mathcal{Z}|), \\ \implies H(g(X_3^k)) &\leq k(R_{31} + R_{32} + \epsilon) \log |\mathcal{A}|,\end{aligned}$$

the second inequality uses the definition of rate, and ϵ can be made small enough because $H_{|\mathcal{Z}|}(\mathbf{Z}_{31}) \leq H_{|\mathcal{Z}|}(X_3^k) = k$ and similarly for $H_{|\mathcal{Z}|}(\mathbf{Z}_{32})$. ■

Accordingly, in the rest of the paper, we focus on the quantities R_1 and R_2 . Using inequality (1) and Lemma 1, we can conclude the following for the sum rate $R_1 + R_2$.

$$\begin{aligned}
\frac{\mathbb{E} \ell(\mathbf{Z}_1) + \mathbb{E} \ell(\mathbf{Z}_2)}{2} &\geq \frac{H_{|\mathcal{Z}|}(\mathbf{Z}_1) + H_{|\mathcal{Z}|}(\mathbf{Z}_2)}{2} - \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_1) + |\mathcal{Z}|) - \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_2) + |\mathcal{Z}|) \\
&\stackrel{(a)}{\geq} (H_{|\mathcal{Z}|}(f(X_1^k, X_2^k, X_3^k)) + H_{|\mathcal{Z}|}(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k))) / 2 \\
&\quad - \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_1) + |\mathcal{Z}|) - \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_2) + |\mathcal{Z}|) \\
&\stackrel{(b)}{\geq} (H_{|\mathcal{Z}|}(f(X_1^k, X_2^k, X_3^k)) + H_{|\mathcal{Z}|}(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k), X_3^k)) / 2 \\
&\quad - k \left(\log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_1) + |\mathcal{Z}|) + \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_2) + |\mathcal{Z}|) \right) / k \\
\implies \frac{R_1 + R_2 + \epsilon}{2} &\geq \frac{\mathbb{E} \ell(\mathbf{Z}_1) + \mathbb{E} \ell(\mathbf{Z}_2)}{2k \log_{|\mathcal{Z}|} |\mathcal{A}|} \\
&\geq \frac{\alpha + H_{|\mathcal{Z}|}(f(X_1^k, X_2^k, X_3^k)) / k}{2 \log_{|\mathcal{Z}|} |\mathcal{A}|} - \epsilon' \text{ for } \epsilon, \epsilon' > 0 \text{ and large } k, \tag{2}
\end{aligned}$$

where inequality (a) is because $H(\mathbf{Z}_1, \mathbf{Z}_2, f(X_1^k, X_2^k, X_3^k)) = H(\mathbf{Z}_1, \mathbf{Z}_2)$ by the zero error criterion, inequality (b) is true as conditioning reduces entropy, and implication (2) is true by the value of α obtained in Section III-B, by the definition of rate, and the fact that $H_{|\mathcal{Z}|}(\mathbf{Z}_1) \leq H(X_1^k, X_3^k) = 2k \log_{|\mathcal{Z}|} |\mathcal{A}|$ and similarly $H_{|\mathcal{Z}|}(\mathbf{Z}_2) \leq H(X_2^k, X_3^k) = 2k \log_{|\mathcal{Z}|} |\mathcal{A}|$.

Focusing on just $H_{|\mathcal{Z}|}(\mathbf{Z}_u)$ for either $u = 1$ or 2 , we have the following inequality.

$$\begin{aligned}
\mathbb{E} \ell(\mathbf{Z}_u) &\geq H_{|\mathcal{Z}|}(\mathbf{Z}_u) - 2 \log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_u) + |\mathcal{Z}|) \\
&\stackrel{(a)}{\geq} H_{|\mathcal{Z}|}(\mathbf{Z}_u | X_3^k) - 2(\log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_u) + |\mathcal{Z}|)) \stackrel{(b)}{\geq} \gamma k - 2k(\log_{|\mathcal{Z}|}(H_{|\mathcal{Z}|}(\mathbf{Z}_u) + |\mathcal{Z}|) / k) \\
\implies R_u + \epsilon &\geq \log_{|\mathcal{A}|} |\mathcal{Z}| \mathbb{E} \ell(\mathbf{Z}_u) / k \geq \gamma / \log_{|\mathcal{Z}|} |\mathcal{A}| - \epsilon' \text{ for large } k, \tag{3}
\end{aligned}$$

where inequality (a) is true because conditioning reduces entropy. A procedure to obtain the value of γ in inequality (b) will be described in the next section (c.f. equation (5)).

A. Lower bound on the conditional entropy

In this section we use the structure of the demand function to obtain a lower bound on the conditional entropy of the descriptions transmitted on the edges (s_3, s_1) and (s_3, s_2) . This enables us to find the values of α and γ used in the previous section. To do this, we define quantities which denote the minimum number of distinct $(\mathbf{Z}_1, \mathbf{Z}_2)$ -labels that allow the terminal to recover $f(X_1^k, X_2^k, X_3^k)$ with zero error. These quantities have been defined generally in [18], [19] and we adapt them to our particular network instance. Let lowercase letters $x_j, y_j \in \mathcal{A}$ denote realizations of the message random variable $X_j, j \in \{1, 2, 3\}$. For ease of notation, the ordering of the arguments of $f(X_1, X_2, X_3)$ is ignored and subscripts indicate which message random variable a particular realization corresponds to.

Definition 3: For two different message realizations $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathcal{A}^3$ such that $x_3 = y_3 \triangleq a_3$, we say² that $x_1 \stackrel{a_3}{\equiv} y_1|_1$ if and only if $f(x_1, x_2, a_3) = f(y_1, y_2, a_3)$ for all $x_2 = y_2 \in \mathcal{A}$. Similarly, for two different message realizations $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathcal{A}^3$ such that $x_3 = y_3 \triangleq a_3$, we say that $x_2 \stackrel{a_3}{\equiv} y_2|_2$ if and only if $f(x_1, x_2, a_3) = f(y_1, y_2, a_3)$ for all $x_1 = y_1 \in \mathcal{A}$.

Note that for both $u = 1, 2$ and any value of $a_3 \triangleq x_3 = y_3$,

- $x_u = y_u$ implies $x_u \stackrel{a_3}{\equiv} y_u|_u$,
- $x_u \stackrel{a_3}{\equiv} y_u|_u$ implies $y_u \stackrel{a_3}{\equiv} x_u|_u$, and
- $x_u \stackrel{a_3}{\equiv} w_u|_u$ and $w_u \stackrel{a_3}{\equiv} y_u|_u$ implies $x_u \stackrel{a_3}{\equiv} y_u|_u$.

Thus $\stackrel{a_3}{\equiv}|_u$ is an equivalence relation on \mathcal{A} for any demand function $f(X_1, X_2, X_3)$, choice of $u \in \{1, 2\}$ and $a_3 \in \mathcal{A}$. The number of equivalence classes of \mathcal{A} induced by $\stackrel{a_3}{\equiv}|_u$ is denoted as $V_u(a_3)$ for both $u = 1$ and 2 .

Illustration 1: Throughout the paper we will use a running example to illustrate the various quantities defined. In the network of Figure 1 we choose the message alphabet to be the finite field of order 3 (denoted $GF(3)$) and the demand function³ values are as listed in Table I. From Table Ib we have that $0 \stackrel{1}{\equiv} 1|_1$ and $1 \stackrel{1}{\equiv} 2|_2$. This is because for any value of X_2 , the entries in the rows corresponding to $X_1 = 0$ and $X_1 = 1$ in Table Ib are the same, giving $0 \stackrel{1}{\equiv} 1|_1$. A similar statement is true for the

²Read as ‘ x_1 is a_3 -equivalent to y_1 for the message at s_1 ’.

³Any function from $GF(3)^3 \rightarrow GF(3)$ can be written as a multivariate polynomial in its arguments. Here it is $X_1^2 X_2^2 X_3 - X_1^2 X_2 X_3^2 + X_1 X_2^2 X_3^2 + X_1^2 X_2 X_3 + X_1 X_2^2 X_3 - X_2^2 X_3^2 - X_1 X_3 - X_2 X_3 - X_3^2 + X_1 - X_2 + X_3$.

TABLE I

FUNCTION TABLE FOR A DEMAND FUNCTION TO BE COMPUTED OVER THE NETWORK IN FIGURE 1. THE MESSAGE ALPHABET IS $\mathcal{A} = GF(3)$. TABLE IA SHOWS THE FUNCTION VALUES FOR ALL (X_1, X_2) PAIRS WHEN $X_3 = 0$, TABLE IB SHOWS THE FUNCTION VALUES WHEN $X_3 = 1$ AND TABLE IC SHOWS THE FUNCTION VALUES WHEN $X_3 = 2$.

$X_3 = 0$	X_2			
	0	1	2	
X_1	0	0	2	1
	1	1	0	2
	2	2	1	0

(a)

$X_3 = 1$	X_2			
	0	1	2	
X_1	0	0	0	0
	1	0	0	0
	2	1	0	0

(b)

$X_3 = 2$	X_2			
	0	1	2	
X_1	0	1	1	0
	1	1	1	1
	2	1	1	1

(c)

columns corresponding to $X_2 = 1$ and $X_2 = 2$ in Table Ib. Since $0 \stackrel{1}{\equiv} 1 \not\equiv 2|_1$, we see that the relation $\stackrel{1}{\equiv} |_1$ partitions the alphabet $\mathcal{A} = GF(3)$ into two equivalence classes i.e. $\{0, 1\}$ and $\{2\}$ and hence $V_1(1) = 2$. From Table I one can verify the following equivalence classes of $GF(3)$ for each kind of relation in the considered demand function.

$$\stackrel{0}{\equiv} |_1 : \{0\} \cup \{1\} \cup \{2\}, \quad \stackrel{1}{\equiv} |_1 : \{0, 1\} \cup \{2\}, \quad \stackrel{2}{\equiv} |_1 : \{0\} \cup \{1, 2\}, \quad (4a)$$

$$\stackrel{0}{\equiv} |_2 : \{0\} \cup \{1\} \cup \{2\}, \quad \stackrel{1}{\equiv} |_2 : \{0\} \cup \{1, 2\}, \quad \stackrel{2}{\equiv} |_2 : \{0, 1\} \cup \{2\}. \quad (4b)$$

Hence for this demand function we have that $V_1(0) = V_2(0) = 3$, $V_1(1) = V_2(1) = 2$ and $V_1(2) = V_2(2) = 2$.

Lemma 3 (Adapted from Lemma 3 in [19]): The source node s_1 is the only source disconnected from the terminal if the edge (s_1, t) is removed from the network, even though s_3 has a path connecting it to the terminal through the edge (s_1, t) . A similar statement also holds for the case of source node s_2 and edge (s_2, t) . For two realizations $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathcal{A}^3$ such that $x_3 = y_3 = a_3$ a valid network code must transmit different \mathbf{Z}_1 -labels on the edge (s_1, t) if $x_1 \not\stackrel{a_3}{\equiv} y_1|_1$ and different \mathbf{Z}_2 -labels on the edge (s_2, t) if $x_2 \not\stackrel{a_3}{\equiv} y_2|_2$.

Thus, for a given realization a_3 of X_3 , what matters is whether the function takes different values for different realizations of X_1 for some value of $X_2 \in \mathcal{A}$. If it does, then those two realizations of X_1 belong to different $\stackrel{a_3}{\equiv} |_1$ equivalence classes and hence by Lemma 3 must have distinct labels transmitted over the edge (s_1, t) . A similar argument can be made for the labels transmitted on the (s_2, t) edge based on the $\stackrel{a_3}{\equiv} |_2$ equivalence class of $X_2 \in \mathcal{A}$. For either $u = 1$ or 2 , the $\stackrel{a_3}{\equiv} |_u$ relation has a natural extension to vector realizations $\mathbf{x}_u \in \mathcal{A}^k$. We then have the following lemma, whose proof is in Appendix B. We use lowercase boldface to denote vectors whose length is inferred from the context henceforth, like $\mathbf{a}_3 \triangleq (a_3^{(1)}, a_3^{(2)}, \dots, a_3^{(k)})$.

Lemma 4: Consider a block of k independent realizations of X_1, X_2 and X_3 and let $\mathbf{a}_3 \in \mathcal{A}^k$ be the realization for X_3 . Then for $u \in \{1, 2\}$, the number of distinct \mathbf{Z}_u -labels that must be transmitted on the edge (s_u, t) to allow the terminal to recover $f(X_1^k, X_2^k, X_3^k)$ with zero error is at least $V_u(\mathbf{a}_3) \triangleq \prod_{i=1}^k V_u(a_3^{(i)})$. Equivalently, let $\stackrel{\mathbf{a}_3}{\equiv} |_u$ denote the collection of equivalence relations of Definition 3 for each component of \mathbf{a}_3 . In this notation, we have that

$$\mathbf{x}_u \stackrel{\mathbf{a}_3}{\equiv} \mathbf{y}_u|_u \Leftrightarrow x_u^{(j)} \stackrel{a_3^{(j)}}{\equiv} y_u^{(j)}|_u, \quad \forall j \in \{1, 2, \dots, k\}.$$

If $x_1 \not\stackrel{a_3}{\equiv} y_1|_1$, then $\phi_{(s_1, t)}(\mathbf{x}_1, \mathbf{a}_3) \neq \phi_{(s_1, t)}(\mathbf{y}_1, \mathbf{a}_3)$, i.e., their \mathbf{Z}_1 labels must be different. An analogous statement can be seen to be true for the \mathbf{Z}_2 label as well.

For either $u = 1$ or 2 and each $i \in \{1, 2, \dots, k\}$, the equivalence classes under $\stackrel{a_3^{(i)}}{\equiv} |_u$ are denoted as $\text{Cl}_u^{(j)}(a_3^{(i)})$, where the superscript $j \in \{1, 2, \dots, V_u(a_3^{(i)})\}$ indexes the classes such that

$$|\text{Cl}_u^{(1)}(a_3^{(i)})| \geq |\text{Cl}_u^{(2)}(a_3^{(i)})| \geq \dots \geq |\text{Cl}_u^{(V_u(a_3^{(i)}))}(a_3^{(i)})|.$$

As described in the proof of Lemma 4, \mathcal{A}^k can be partitioned into $V_u(\mathbf{a}_3) = \prod_{i=1}^k V_u(a_3^{(i)})$ partitions based on the value of \mathbf{a}_3 for each component. Thus the partitions of \mathcal{A}^k under $\stackrel{\mathbf{a}_3}{\equiv} |_u$ can be represented using an index vector \mathbf{v} having k components, each of which satisfies $v^{(i)} \in \{1, 2, \dots, V_u(a_3^{(i)})\}$ and

$$\mathbf{x}_u \in \text{Cl}_u^{(\mathbf{v})}(\mathbf{a}_3) \Leftrightarrow x_u^{(i)} \in \text{Cl}_u^{(v^{(i)})}(a_3^{(i)}), \quad \forall i \in \{1, 2, \dots, k\}.$$

Similar to the scalar case, we add a subscript $t \in \{1, 2, \dots, V_u(\mathbf{a}_3)\}$ to get the index vector \mathbf{v}_t such that the equivalence classes under $\stackrel{\mathbf{a}_3}{\equiv} |_u$ satisfy

$$|\text{Cl}_u^{(\mathbf{v}_t)}(\mathbf{a}_3)| \geq |\text{Cl}_u^{(\mathbf{v}_{t-1})}(\mathbf{a}_3)| \geq \dots \geq |\text{Cl}_u^{(\mathbf{v}_{V_u(\mathbf{a}_3)})}(\mathbf{a}_3)|.$$

From the definition, for every $t \in \{1, 2, \dots, V_u(\mathbf{a}_3)\}$, we have that $\text{Cl}_u^{(\mathbf{v}_t)}(\mathbf{a}_3) = \times_{i=1}^k \text{Cl}_u^{(v_t^{(i)})}(a_3^{(i)})$, i.e., a cartesian product of k scalar equivalence classes and accordingly $|\text{Cl}_u^{(\mathbf{v}_t)}(\mathbf{a}_3)| = \prod_{i=1}^k |\text{Cl}_u^{(v_t^{(i)})}(a_3^{(i)})|$.

The following notation is useful in characterizing all valid probability mass functions for \mathbf{Z}_u for both $u = 1, 2$ and also the pair label $(\mathbf{Z}_1, \mathbf{Z}_2)$. For any index i of a vector \mathbf{p} , we use $p^{[i]}$ to denote the i th component of \mathbf{p} when it is arranged in non-increasing order. For two vectors \mathbf{p}, \mathbf{q} of the same length l , the vector \mathbf{p} is majorized by \mathbf{q} , denoted as $\mathbf{p} \prec \mathbf{q}$, if the following holds.

$$\begin{aligned} \sum_{i=1}^t p^{[i]} &\leq \sum_{i=1}^t q^{[i]}, \text{ for all } t = 1, 2, \dots, l-1, \text{ and} \\ \sum_{i=1}^l p^{[i]} &= \sum_{i=1}^l q^{[i]}. \end{aligned}$$

As an example, the vector $[0.5 \ 0.5]$ is majorized by $[0.25 \ 0.75]$. Note that any vector \mathbf{p} is majorized by itself.

Lemma 4 gives a lower bound on the number of distinct \mathbf{Z}_1 and \mathbf{Z}_2 labels that must be used by a source-network code for a particular realization of X_3^k . Using this and the uniform i.i.d. assumption on the messages, we can characterize the set of valid p.m.f.s for the conditional probability of the \mathbf{Z}_u label given a realization \mathbf{a}_3 of the message X_3^k .

Lemma 5: Let $\mathbb{Z}_{\geq 0}$ denote the set of non-negative integers and superscript \top indicates the transpose operation. Consider the partitions of \mathcal{A}^k induced by the set of relations $\stackrel{\mathbf{a}_3}{\equiv} |_u$ where \mathbf{a}_3 is a realization of X_3^k and $u = 1$ or 2 . Let $L_u(\mathbf{a}_3) \geq V_u(\mathbf{a}_3)$ be the number of distinct \mathbf{Z}_u labels assigned by an encoding scheme to the $|\mathcal{A}|^k$ different (X_u^k, \mathbf{a}_3) pairs. Define a vector $\mathbf{d}_u(\mathbf{a}_3) \in \mathbb{Z}_{\geq 0}^{L_u(\mathbf{a}_3)}$ as

$$\mathbf{d}_u(\mathbf{a}_3) \triangleq \left[|\text{Cl}_u^{(v_1)}(\mathbf{a}_3)| \quad |\text{Cl}_u^{(v_2)}(\mathbf{a}_3)| \quad \dots \quad |\text{Cl}_u^{(v_{L_u(\mathbf{a}_3)})}(\mathbf{a}_3)| \quad \mathbf{0}_{L_u(\mathbf{a}_3) - V_u(\mathbf{a}_3)} \right]^\top,$$

where $\mathbf{0}_{L_u(\mathbf{a}_3) - V_u(\mathbf{a}_3)}$ indicates a zero vector of length $L_u(\mathbf{a}_3) - V_u(\mathbf{a}_3)$. Then any valid conditional p.m.f. $\mathbf{p} \in \mathbb{R}^{L_u(\mathbf{a}_3)}$ for \mathbf{Z}_u given the value of $X_3^k = \mathbf{a}_3$ satisfies $\mathbf{p} \prec \mathbf{d}_u(\mathbf{a}_3)/|\mathcal{A}|^k$.

Proof: We first note that $\mathbf{d}_u(\mathbf{a}_3)/|\mathcal{A}|^k$ is a valid p.m.f. as its components are non-negative and sum up to 1. Suppose that there is an encoding scheme for \mathbf{Z}_u such that $\Pr\{\mathbf{Z}_u | X_3^k = \mathbf{a}_3\} \triangleq \mathbf{p} \not\prec \mathbf{d}_u(\mathbf{a}_3)/|\mathcal{A}|^k$. Furthermore let \mathbf{p} be supported on $L_u(\mathbf{a}_3)$ components. Then the assumption implies that there is a $t < L_u(\mathbf{a}_3)$ such that

$$\sum_{j=1}^t p^{[j]} > \frac{1}{|\mathcal{A}|^k} \sum_{j=1}^t d_u^{[j]} = \frac{1}{|\mathcal{A}|^k} \sum_{j=1}^t |\text{Cl}_u^{(v_j)}(\mathbf{a}_3)|.$$

Since each realization of X_u^k is equally likely, the RHS in the above equation is the conditional probability given the value of $X_3^k = \mathbf{a}_3$ of the event that X_u^k belongs to one of the t largest equivalence classes under $\stackrel{\mathbf{a}_3}{\equiv} |_u$. The LHS is the conditional probability of observing any of the t most probable \mathbf{Z}_u labels. Hence the encoding scheme gives a total of t distinct \mathbf{Z}_u labels to at least as many (X_u^k, \mathbf{a}_3) pairs for which X_u^k belongs to $t+1$ different equivalence classes under $\stackrel{\mathbf{a}_3}{\equiv} |_u$. By the pigeonhole principle, this contradicts lemma 4. \blacksquare

In order to obtain a lower bound on $H(\mathbf{Z}_u | X_3^k = \mathbf{a}_3)$, we use the order-preserving property of the entropy function with respect to the majorization relation between two vectors [20]. The entropy function $H : \mathbb{R}^{L_u(\mathbf{a}_3)} \rightarrow \mathbb{R}$ is a strictly Schur-concave function [20, Chap. 3], i.e., for two p.m.f.s $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{L_u(\mathbf{a}_3)}$ that are not equal to each other under any permutation of their components, we have that

$$\mathbf{p} \prec \mathbf{q} \implies H(\mathbf{p}) > H(\mathbf{q}).$$

Thus, from lemma 5, we have that $H_{|\mathcal{Z}|}(\mathbf{Z}_u | X_3^k = \mathbf{a}_3) \geq H_{|\mathcal{Z}|}(\mathbf{d}_u(\mathbf{a}_3)/|\mathcal{A}|^k)$. The value of γ as used in equation (3) can be found as follows.

$$\gamma \triangleq \frac{1}{k} \sum_{\mathbf{a}_3 \in \mathcal{A}^k} \Pr(X_3^k = \mathbf{a}_3) H_{|\mathcal{Z}|}(\mathbf{d}_u(\mathbf{a}_3)/|\mathcal{A}|^k) \leq \frac{1}{k} \sum_{\mathbf{a}_3 \in \mathcal{A}^k} \Pr(X_3^k = \mathbf{a}_3) H_{|\mathcal{Z}|}(\mathbf{Z}_u | X_3^k = \mathbf{a}_3) = \frac{H_{|\mathcal{Z}|}(\mathbf{Z}_u | X_3^k)}{k}. \quad (5)$$

Illustration 2: We evaluate the vector $\mathbf{d}_1(\mathbf{a}_3)$ for the example function and consequently obtain a lower bound for $H(\mathbf{Z}_1 | X_3^k)$ and R_1 in this case. As shown in the previous illustration, $V_1(0) = 3$ and $V_1(1) = V_1(2) = 2$. Suppose X_3^k takes the value \mathbf{a}_3 , where \mathbf{a}_3 has m_t components with value t for $t \in \{0, 1, 2\}$ such that $m_0 + m_1 + m_2 = k$. For each component $a_3^{(i)} \in \mathcal{A}, i \in \{1, 2, \dots, k\}$ the scalar equivalence classes under $\stackrel{\mathbf{a}_3^{(i)}}{\equiv} |_u$ are given in equations (4a), (4b), and are represented as below.

$$\begin{aligned} \text{Cl}_u^{(1)}(0) &= \{0\}, \quad \text{Cl}_u^{(2)}(0) = \{1\}, \quad \text{Cl}_u^{(3)}(0) = \{2\}, \quad \text{for both } u = 1, 2 \text{ and} \\ \text{Cl}_1^{(1)}(1) &= \{0, 1\}, \quad \text{Cl}_1^{(2)}(1) = \{2\}, \quad \text{Cl}_1^{(1)}(2) = \{1, 2\}, \quad \text{Cl}_1^{(2)}(2) = \{0\}, \\ \text{Cl}_2^{(1)}(1) &= \{1, 2\}, \quad \text{Cl}_2^{(2)}(1) = \{0\}, \quad \text{Cl}_2^{(1)}(2) = \{0, 1\}, \quad \text{Cl}_2^{(2)}(2) = \{2\}. \end{aligned}$$

Accordingly, the total number of partitions

$$V_1(\mathbf{a}_3) = \prod_{i=1}^k V_1(a_3^{(i)}) = \left(\prod_{i:a_3^{(i)}=0} V_1(0) \right) \left(\prod_{i:a_3^{(i)}=1} V_1(1) \right) \left(\prod_{i:a_3^{(i)}=2} V_1(2) \right) = 3^{m_0} 2^{m_1+m_2}.$$

To find the value of $|\text{Cl}_1^{(v_1)}(\mathbf{a}_3)|$, we pick the scalar partitions for each component that have the largest number of elements under $\stackrel{0}{\equiv} |$, $\stackrel{1}{\equiv} |$ and $\stackrel{2}{\equiv} |$. From the above equations we get $|\text{Cl}_1^{(v_1^{(i)})}(a_3^{(i)})| = 2$ if $a_3^{(i)} \in \{1, 2\}$ and $|\text{Cl}_1^{(v_1^{(i)})}(a_3^{(i)})| = 1$ if $a_3^{(i)} = 0$. Thus $|\text{Cl}_1^{(v_1)}(\mathbf{a}_3)| = 2^{m_1+m_2}$. Furthermore, since there are 3 different largest equivalence classes under $\stackrel{0}{\equiv} |$, we get that

$$|\text{Cl}_1^{(v_1)}(\mathbf{a}_3)| = |\text{Cl}_1^{(v_2)}(\mathbf{a}_3)| = \dots = |\text{Cl}_1^{(v_{3^{m_0}})}(\mathbf{a}_3)| = 2^{m_1+m_2}.$$

Now, we find the number of realizations \mathbf{x}_1 whose components do not necessarily belong to the largest scalar equivalence class at each component. If $t \leq m_1 + m_2$ components of \mathbf{x}_1 are such that either $x_1^{(i)} = 0 \in \text{Cl}_1^{(2)}(a_3^{(i)})$ if $a_3^{(i)} = 2$ or $x_1^{(i)} = 2 \in \text{Cl}_1^{(2)}(a_3^{(i)})$ if $a_3^{(i)} = 1$, then the total number of X_1^k realizations that belong to the same equivalence class as \mathbf{x}_1 is $2^{m_1+m_2-t}$. Thus we have that

$$\begin{aligned} \text{for } t = 1: & \quad |\text{Cl}_1^{(v_{3^{m_0+1}})}(\mathbf{a}_3)| = \dots = |\text{Cl}_1^{(v_{3^{m_0+(m_1+m_2)3^{m_0}}})}(\mathbf{a}_3)| = 2^{m_1+m_2-1}, \\ \text{for } t = 2: & \quad |\text{Cl}_1^{(v_{3^{m_0+(m_1+m_2)3^{m_0+1}}})}(\mathbf{a}_3)| = \dots = |\text{Cl}_1^{(v_{3^{m_0+(m_1+m_2)3^{m_0+(m_1+m_2)3^{m_0}}})}(\mathbf{a}_3)| = 2^{m_1+m_2-2}, \\ & \quad \vdots \\ \text{for } t = m_1 + m_2: & \quad |\text{Cl}_1^{(v_{3^{m_0 \sum_{l=0}^{m_1+m_2-1} (m_1+m_2)^l})}(\mathbf{a}_3)| = \dots = |\text{Cl}_1^{(v_{3^{m_0 \sum_{l=0}^{m_1+m_2} (m_1+m_2)^l})}(\mathbf{a}_3)| = 2^0 = 1. \end{aligned}$$

The above equations determine the components of the vector $\mathbf{d}_1(\mathbf{a}_3)$. We can then evaluate $H_{|\mathcal{Z}|}(\mathbf{d}_1(\mathbf{a}_3)/3^k)$ as

$$\begin{aligned} & 3^{m_0} \cdot \frac{1}{3^{m_0}} \left(\frac{2}{3}\right)^{m_1+m_2} \log_{|\mathcal{Z}|} \frac{3^k}{2^{m_1+m_2}} + 3^{m_0} \binom{m_1+m_2}{1} \cdot \frac{1}{2 \cdot 3^{m_0}} \left(\frac{2}{3}\right)^{m_1+m_2} \log_{|\mathcal{Z}|} \frac{3^k}{2^{m_1+m_2-1}} \\ & + 3^{m_0} \binom{m_1+m_2}{2} \cdot \frac{1}{2^2 \cdot 3^{m_0}} \left(\frac{2}{3}\right)^{m_1+m_2} \log_{|\mathcal{Z}|} \frac{3^k}{2^{m_1+m_2-2}} + \dots + 3^{m_0} \binom{m_1+m_2}{m_1+m_2} \cdot \frac{1}{2^{m_1+m_2} \cdot 3^{m_0}} \left(\frac{2}{3}\right)^{m_1+m_2} \log_{|\mathcal{Z}|} \frac{3^k}{2^0} \\ & = (k \log_{|\mathcal{Z}|} 3 - (m_1+m_2) \log_{|\mathcal{Z}|} 2) \left(\frac{2}{3}\right)^{m_1+m_2} \left(1 + \binom{m_1+m_2}{1} 2^{-1} + \dots + \binom{m_1+m_2}{m_1+m_2} 2^{-m_1-m_2}\right) \\ & + \log_{|\mathcal{Z}|} 2 \left(\frac{2}{3}\right)^{m_1+m_2} \left(\binom{m_1+m_2}{1} 2^{-1} + 2 \binom{m_1+m_2}{2} 2^{-2} + \dots + (m_1+m_2) \binom{m_1+m_2}{m_1+m_2} 2^{-m_1-m_2}\right) \\ & = (k \log_{|\mathcal{Z}|} 3 - (m_1+m_2) \log_{|\mathcal{Z}|} 2) \left(\frac{2}{3}\right)^{m_1+m_2} \left(\frac{3}{2}\right)^{m_1+m_2} + \log_{|\mathcal{Z}|} 2 \left(\frac{2}{3}\right)^{m_1+m_2} \frac{m_1+m_2}{3} \left(\frac{3}{2}\right)^{m_1+m_2} \\ & = k \log_{|\mathcal{Z}|} 3 - \frac{2(m_1+m_2)}{3} \log_{|\mathcal{Z}|} 2. \end{aligned}$$

It follows that

$$\begin{aligned} H_{|\mathcal{Z}|}(\mathbf{Z}_1|X_3^k) &= \sum_{\mathbf{a}_3} \Pr\{X_3^k = \mathbf{a}_3\} H_{|\mathcal{Z}|}(\mathbf{Z}_1|X_3^k = \mathbf{a}_3) = \sum_{\substack{m_1, m_2=0, \\ m_1+m_2 \leq k}}^k \sum_{\substack{\mathbf{a}_3 \text{ has } m_1 \text{ 1's,} \\ m_2 \text{ 2's}}} \Pr\{X_3^k = \mathbf{a}_3\} H_{|\mathcal{Z}|}(\mathbf{Z}_1|X_3^k = \mathbf{a}_3) \\ &\geq \sum_{\substack{m_1, m_2=0, \\ m_1+m_2 \leq k}}^k \frac{k!}{3^k m_1! m_2! (k-m_1-m_2)!} \left(k \log_{|\mathcal{Z}|} 3 - \frac{2(m_1+m_2)}{3} \log_{|\mathcal{Z}|} 2\right) \\ &= k \log_{|\mathcal{Z}|} 3 - \frac{2 \log_{|\mathcal{Z}|} 2}{3} \frac{2k3^{k-1}}{3^k} = k(\log_{|\mathcal{Z}|} 3 - \frac{4}{9} \log_{|\mathcal{Z}|} 2). \end{aligned}$$

Thus the value of γ is $(\log_{|\mathcal{Z}|} 3 - \frac{4}{9} \log_{|\mathcal{Z}|} 2)$ and from equation (3), $R_1 + \epsilon \geq \log_{|\mathcal{A}|} 3 - \frac{4}{9} \log_{|\mathcal{A}|} 2 \approx 0.7196$.

To obtain a lower bound on the conditional entropy of the pair of labels as in equation (2), we characterize the family of valid conditional p.m.f.s for the pair $(\mathbf{Z}_1, \mathbf{Z}_2)$ given the values of the demand function $f(X_1^k, X_2^k, X_3^k)$ and the realization of the message X_3^k . Towards this end we first find the number of distinct $(\mathbf{Z}_1, \mathbf{Z}_2)$ -labels that must be assigned by the network code to message tuples that result in a particular value, say, $\mathbf{b} \in \mathcal{B}^k$ of the demand function $f(X_1^k, X_2^k, X_3^k)$. The set $A_3(\mathbf{b})$ has all possible realizations \mathbf{a}_3 of X_3^k that can result in the value of \mathbf{b} for the demand function, i.e.,

$$A_3(\mathbf{b}) \triangleq \{\mathbf{a}_3 \in \mathcal{A}^k : \exists \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}^k \text{ such that } f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = \mathbf{b}\}.$$

Let $M(\mathbf{a}_3, \mathbf{b})$ denote the number of distinct $(\mathbf{Z}_1, \mathbf{Z}_2)$ pair labels used for message tuples that have $X_3^k = \mathbf{a}_3$ and $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = \mathbf{b}$. Consider two message tuples $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3)$ and $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{a}_3)$ which satisfy $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = f(\mathbf{y}_1, \mathbf{y}_2, \mathbf{a}_3) = \mathbf{b}$.

TABLE II
THE SETS $\mathcal{V}_{12}(a_3, b)$ FOR DIFFERENT VALUES OF a_3 AND b IN DIFFERENT ROWS AND COLUMNS RESPECTIVELY.

a_3	$b = 0$	$b = 1$	$b = 2$
0	$\{(Cl_1^{(1)}, Cl_2^{(1)}), (Cl_1^{(2)}, Cl_2^{(2)}), (Cl_1^{(3)}, Cl_2^{(3)})\}$	$\{(Cl_1^{(1)}, Cl_2^{(3)}), (Cl_1^{(2)}, Cl_2^{(1)}), (Cl_1^{(3)}, Cl_2^{(2)})\}$	$\{(Cl_1^{(1)}, Cl_2^{(2)}), (Cl_1^{(2)}, Cl_2^{(3)}), (Cl_1^{(3)}, Cl_2^{(1)})\}$
1	$\{(Cl_1^{(1)}, Cl_2^{(1)}), (Cl_1^{(1)}, Cl_2^{(2)}), (Cl_1^{(2)}, Cl_2^{(1)})\}$	$\{(Cl_1^{(2)}, Cl_2^{(2)})\}$	\emptyset
2	$\{(Cl_1^{(2)}, Cl_2^{(2)})\}$	$\{(Cl_1^{(1)}, Cl_1^{(1)}), (Cl_1^{(1)}, Cl_2^{(2)}), (Cl_1^{(2)}, Cl_2^{(1)})\}$	\emptyset

If either $\mathbf{x}_1 \not\equiv \mathbf{y}_1|_1$ or $\mathbf{x}_2 \not\equiv \mathbf{y}_2|_2$, then the pair of labels $(\mathbf{Z}_1, \mathbf{Z}_2)$ assigned to the two message tuples must be different. This motivates us to define the pair index set:

$$\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b}) \triangleq \left\{ (Cl_1^{(v_j)}, Cl_2^{(w_t)}) : \begin{array}{l} \exists \mathbf{x}_1 \in Cl_1^{(v_j)}(\mathbf{a}_3), \mathbf{x}_2 \in Cl_2^{(w_t)}(\mathbf{a}_3) \text{ s.t. } f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = \mathbf{b}, \\ 1 \leq j \leq V_1(\mathbf{a}_3), 1 \leq t \leq V_2(\mathbf{a}_3) \end{array} \right\}.$$

The above discussion then implies that $M(\mathbf{a}_3, \mathbf{b}) \geq |\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|$. By the definition, if $(Cl_1^{(v)}, Cl_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$ then for every $i \in \{1, 2, \dots, k\}$, the pair of scalar equivalence classes $(Cl_1^{(v^{(i)})}, Cl_2^{(w^{(i)})}) \in \mathcal{V}_{12}(a_3^{(i)}, b^{(i)})$.

Illustration 3: Consider block size $k = 1$ and a realization $b = 0$ of the demand function of Table I. In the previous illustration we evaluated that $Cl_1^{(1)}(1) = \{0, 1\}$, $Cl_1^{(2)}(1) = \{2\}$ and $Cl_2^{(1)}(1) = \{1, 2\}$, $Cl_2^{(2)}(1) = \{0\}$. Then we can evaluate that the pair index set $\mathcal{V}_{12}(1, 0) = \{(Cl_1^{(1)}, Cl_2^{(1)}), (Cl_1^{(1)}, Cl_2^{(2)}), (Cl_1^{(2)}, Cl_2^{(1)})\}$. Note that the pair $(Cl_1^{(2)}, Cl_2^{(2)}) \notin \mathcal{V}_{12}(1, 0)$ as the elements of that pair of equivalence classes do not result in the demand function value of 0, i.e.,

$$Cl_1^{(2)}(1) = \{2\}, Cl_2^{(2)}(1) = \{0\} \text{ but for } x_1 = 2, x_2 = 0, x_3 = 1, f(x_1, x_2, x_3) = 1 \neq 0.$$

Thus in this case we have that $|\mathcal{V}_{12}(1, 0)| = 3$. The other pair index sets are summarized in Table II.

We now explicitly derive a p.m.f. whose entropy is a lower bound to the conditional entropy $H(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}, X_3^k = \mathbf{a}_3)$. Let $A_{123}(\mathbf{b}) \subseteq \mathcal{A}^k \times \mathcal{A}^k \times \mathcal{A}^k$ contain all message tuples that are present in the pre-image of the demand function value of \mathbf{b} . We also let $A_{123}(\mathbf{b}, \mathbf{a}_3) \subseteq A_{123}(\mathbf{b})$ denote the subset of message tuples for which the realization of X_3^k is \mathbf{a}_3 . Suppose that $(Cl_1^{(v)}, Cl_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$. The number of different message tuples that cause the membership of the equivalence class pair $(Cl_1^{(v)}, Cl_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$ is denoted as follows.

$$h_{\mathbf{a}_3}(\mathbf{v}, \mathbf{w}) \triangleq \left| \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 \in Cl_1^{(v)}(\mathbf{a}_3), \mathbf{x}_2 \in Cl_2^{(w)}(\mathbf{a}_3), f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = \mathbf{b}\} \right|, \quad (6)$$

$$= |Cl_1^{(v)}(\mathbf{a}_3)| \cdot |Cl_2^{(w)}(\mathbf{a}_3)|, \quad (7)$$

$$= \prod_{i=1}^k |Cl_1^{(v^{(i)})}(a_3^{(i)})| \cdot |Cl_2^{(w^{(i)})}(a_3^{(i)})| = \prod_{i=1}^k h_{a_3^{(i)}}(v^{(i)}, w^{(i)}). \quad (8)$$

Equality (7) is true above as by Definition 3 every element of an equivalence class under $\stackrel{a_3}{\equiv}|_1$ results in the same demand function value (while the other message \mathbf{x}_2 is held constant), and since $(Cl_1^{(v)}, Cl_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$, there is at least one $\mathbf{x}_1 \in Cl_1^{(v)}(\mathbf{a}_3)$ and one $\mathbf{x}_2 \in Cl_2^{(w)}(\mathbf{a}_3)$ such that $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{a}_3) = \mathbf{b}$. Hence every other pair of elements in $Cl_1^{(v)}(\mathbf{a}_3) \times Cl_2^{(w)}(\mathbf{a}_3)$ would also result in the same demand function value with $X_3^k = \mathbf{a}_3$.

Illustration 4: For block size $k = 1$ and demand function realization $b = 0$, we can check that $A_3(0) = \{0, 1, 2\}$. Following the indexing of the equivalence partitions and Table II we have that $h_1(1, 1) = |Cl_1^{(1)}(1)| \cdot |Cl_2^{(1)}(1)| = 4$ as $Cl_1^{(1)}(1) = \{0, 1\}$ and $Cl_2^{(1)}(1) = \{1, 2\}$. One can similarly check that $h_1(1, 2) = h_1(2, 1) = 2$ and for other values of a_3 , that $h_0(1, 1) = h_0(2, 2) = h_0(3, 3) = 1$ and $h_2(2, 2) = 1$.

For block size $k = 3$ and $\mathbf{b} = (1, 2, 1)$ we can check that $\mathbf{a}_3 = (0, 0, 1) \in A_3(\mathbf{b})$. Then the equivalence class pairs under \mathbf{a}_3 that result in this \mathbf{b} can be obtained in the following manner. If $(Cl_1^{(v)}, Cl_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$, then for each component, using Table II we have that

- $(Cl_1^{(v^{(1)})}, Cl_2^{(w^{(1)})}) \in \mathcal{V}_{12}(a_3^{(1)}, b^{(1)}) = \mathcal{V}_{12}(0, 1) = \{(Cl_1^{(1)}, Cl_2^{(3)}), (Cl_1^{(2)}, Cl_2^{(1)}), (Cl_1^{(3)}, Cl_2^{(2)})\}$,
- $(Cl_1^{(v^{(2)})}, Cl_2^{(w^{(2)})}) \in \mathcal{V}_{12}(a_3^{(2)}, b^{(2)}) = \mathcal{V}_{12}(0, 2) = \{(Cl_1^{(1)}, Cl_2^{(2)}), (Cl_1^{(2)}, Cl_2^{(3)}), (Cl_1^{(3)}, Cl_2^{(1)})\}$,
- $(Cl_1^{(v^{(3)})}, Cl_2^{(w^{(3)})}) \in \mathcal{V}_{12}(a_3^{(3)}, b^{(3)}) = \mathcal{V}_{12}(1, 1) = \{(Cl_1^{(2)}, Cl_2^{(2)})\}$.

As $(Cl_1^{(v)}, Cl_2^{(w)}) = (\times_{i=1}^3 Cl_1^{(v^{(i)})}, \times_{i=1}^3 Cl_2^{(w^{(i)})})$, we get that $|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})| = 9$. One of these nine equivalence class pairs is the pair $(Cl_1^{((1,1,2))}, Cl_2^{((3,2,2))})$. Then using equation (8) we have that

$$h_{(0,0,1)}((1, 1, 2), (3, 2, 2)) = h_0(1, 3) \cdot h_0(1, 2) \cdot h_1(2, 2) = 1 \cdot 1 \cdot 1 = 1.$$

The proof of the following lemma is similar in spirit to that of Lemma 5.

Lemma 6: For any $\mathbf{a}_3 \in A_3(\mathbf{b})$, define the vector

$$\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} \triangleq [h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1) \quad h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_2) \quad \cdots \quad h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_{|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|}) \quad \mathbf{0}_{M(\mathbf{a}_3, \mathbf{b}) - |\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|}]^T,$$

where $\mathbf{0}_t$ indicates a vector of zeros of length t and subscript j in $(\mathbf{v}, \mathbf{w})_j$ indexes all the equivalence class pairs such that

$$h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1) \geq h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_2) \geq \dots \geq h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_{|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|}).$$

Then all conditional probability mass functions $\mathbf{p} \in \mathbb{R}_{\geq 0}^{M(\mathbf{a}_3, \mathbf{b})}$ on $M(\mathbf{a}_3, \mathbf{b})$ valid $(\mathbf{z}_1, \mathbf{z}_2)$ -labels given the value \mathbf{b} of the demand function and the realization \mathbf{a}_3 of X_3^k satisfy $\mathbf{p} \prec \mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|$.

It will be useful for analysis to define the following index sets of a demand function realization $\mathbf{b} \in \mathcal{B}^k$ and the message realization $\mathbf{a}_3 \in A_3(\mathbf{b})$.

Definition 4: For every $p \in \mathcal{B}$, let $\mathcal{I}_p(\mathbf{b}) \subseteq \{1, 2, \dots, k\}$ be the index set of components of the demand function realization \mathbf{b} that are equal to p , i.e.,

$$\mathcal{I}_p(\mathbf{b}) = \{i : b^{(i)} = p, i \in \{1, 2, \dots, k\}\}.$$

We can similarly define the index set $\mathcal{J}_q(\mathbf{a}_3)$ for every $q \in \mathcal{A}$.

Illustration 5: We evaluate $H(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|)$ for a general block length k of the example demand function. To specify $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$, we need to find $|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|$ and $h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_j)$ for all $j \in \{1, 2, \dots, |\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|\}$. Suppose the realization $\mathbf{b} \in \{0, 1, 2\}^k$ has m_1 1's, m_2 2's and $k - m_1 - m_2$ 0's, and the realization of X_3^k is some $\mathbf{a}_3 \in A_3(\mathbf{b})$. Let $t_{p,q}$ for any $p \in \mathcal{B}, q \in \mathcal{A}$ be defined as $t_{p,q} \triangleq |\mathcal{I}_p(\mathbf{b}) \cap \mathcal{J}_q(\mathbf{a}_3)|$. Note that $t_{2,1} = t_{2,2} = 0$ for any choice of \mathbf{b} and \mathbf{a}_3 .

a) $|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})|$: We find the number of non-zero components of $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$, i.e. $|\mathcal{V}_{12}(\mathbf{b}, \mathbf{a}_3)|$, as follows. For every $(\text{Cl}_1^{(v)}, \text{Cl}_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$, we have from Table II that for the index i ,

- if $i \in (\mathcal{I}_0(\mathbf{b}) \cap \mathcal{J}_2(\mathbf{a}_3)) \cup (\mathcal{I}_1(\mathbf{b}) \cap \mathcal{J}_1(\mathbf{a}_3))$, there is only one possible choice for the scalar equivalence class pair $(\text{Cl}_1^{(v^{(i)})}, \text{Cl}_2^{(w^{(i)})})$,
- else if i is not in the previous index set, then there are three possible choices for the scalar equivalence class pair $(\text{Cl}_1^{(v^{(i)})}, \text{Cl}_2^{(w^{(i)})})$.

Thus the total number of equivalence class pairs for the choice of \mathbf{b} and \mathbf{a}_3 is $|\mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})| = 3^{k-t_{0,2}-t_{1,1}}$. Next we evaluate the components of the vector $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$.

b) $h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1)$: Based on the subscript indexing of the pair (\mathbf{v}, \mathbf{w}) and equation (8), the value of $h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1)$ is obtained by counting the number of message tuples in the equivalence class pair that has the largest scalar equivalence class pair for each component. As evaluated in the previous illustration, the largest equivalence class pair for every $i \in \mathcal{I}_0(\mathbf{b}) \cap \mathcal{J}_1(\mathbf{a}_3)$ has $h_1(1, 1) = 4$ message tuples, and for every $i \in \mathcal{I}_1(\mathbf{b}) \cap \mathcal{J}_2(\mathbf{a}_3)$ the largest equivalence class again has $h_2(1, 1) = 4$ message tuples. For all other values of i , the largest equivalence class has a single input tuple. Hence we have that $h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1) = 4^{t_{0,1}+t_{1,2}}$. Note that there are three different choices for the largest scalar equivalence class when i belongs to one of the following sets.

- $i \in \mathcal{I}_0(\mathbf{b}) \cap \mathcal{J}_0(\mathbf{a}_3)$: there are $t_{0,0} = k - m_1 - m_2 - t_{0,1} - t_{0,2}$ such components,
- $i \in \mathcal{I}_1(\mathbf{b}) \cap \mathcal{J}_0(\mathbf{a}_3)$: there are $t_{1,0} = m_1 - t_{1,1} - t_{1,2}$ such components, and
- $i \in \mathcal{I}_2(\mathbf{b}) \cap \mathcal{J}_0(\mathbf{a}_3)$: there are $t_{2,0} = m_2$ such components.

Thus there are $3^{k-m_1-m_2-t_{0,1}-t_{0,2}} \cdot 3^{m_1-t_{1,1}-t_{1,2}} \cdot 3^{m_2} = 3^{k-t_{0,1}-t_{0,2}-t_{1,1}-t_{1,2}}$ components of $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$ which have the same value. Let $k' \triangleq k - t_{0,1} - t_{0,2} - t_{1,1} - t_{1,2}$. Hence we have that

$$h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_1) = h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_2) = \dots = h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_{3^{k'}}) = 4^{t_{0,1}+t_{1,2}}.$$

c) $h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_c)$ for $c > 3^{k'}$: Next we consider the case when not every component of a message tuple $(\mathbf{x}_1, \mathbf{x}_2) \in \text{Cl}_1^{(v)} \times \text{Cl}_2^{(w)}$ is present in the largest scalar equivalence class pair. Suppose that $(\text{Cl}_1^{(v)}, \text{Cl}_2^{(w)})$ is such that

- at u_0 indices from $\mathcal{I}_0(\mathbf{b}) \cup \mathcal{J}_1(\mathbf{a}_3)$, the equivalence class pair is either $(\text{Cl}_1^{(1)}, \text{Cl}_2^{(2)})$ or $(\text{Cl}_1^{(2)}, \text{Cl}_2^{(1)})$, and
- at u_1 indices from $\mathcal{I}_1(\mathbf{b}) \cup \mathcal{J}_2(\mathbf{a}_3)$, the equivalence class pair is either $(\text{Cl}_1^{(1)}, \text{Cl}_2^{(2)})$ or $(\text{Cl}_1^{(2)}, \text{Cl}_2^{(1)})$.

Let $c > 3^{k'}$ be the index in $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$ corresponding to this equivalence class pair. We have that $h_1(1, 2) = h_1(2, 1) = 2$ and $h_2(1, 2) = h_2(2, 1) = 2$. Then we get that

$$h_{\mathbf{b}, \mathbf{a}_3}^{(c)} = h_{\mathbf{a}_3}((\mathbf{v}, \mathbf{w})_c) = 4^{t_{0,1}-u_0+t_{1,2}-u_1} \cdot 2^{u_0+u_1} = \frac{4^{t_{0,1}+t_{1,2}}}{2^{u_0+u_1}}.$$

Thus the number of components of $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$ that have the same value as $h_{\mathbf{b}, \mathbf{a}_3}^{(c)}$ are

$$\binom{t_{0,1}}{u_0} 2^{u_0} \binom{t_{1,2}}{u_1} 2^{u_1} 3^{k'}.$$

Thus, the vector $\mathbf{h}_{\mathbf{b}, \mathbf{a}_3}$ is as follows, where u_0 and u_1 are indices satisfying $1 \leq u_0 \leq t_{0,1}$ and $1 \leq u_1 \leq t_{1,2}$.

$$\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} = \left[\underbrace{4^{t_{0,1}+t_{1,2}} \dots 4^{t_{0,1}+t_{1,2}}}_{3^{k'}} \dots \underbrace{\frac{4^{t_{0,1}+t_{1,2}}}{2^{u_0+u_1}} \dots \frac{4^{t_{0,1}+t_{1,2}}}{2^{u_0+u_1}}}_{\binom{t_{0,1}}{u_0} 2^{u_0} \binom{t_{1,2}}{u_1} 2^{u_1} 3^{k'}} \dots \underbrace{\frac{4^{t_{0,1}+t_{1,2}}}{2^{t_{0,1}+t_{1,2}}} \dots \frac{4^{t_{0,1}+t_{1,2}}}{2^{t_{0,1}+t_{1,2}}}}_{2^{t_{0,1}} 2^{t_{1,2}} 3^{k'}} \right]^\top. \quad (9)$$

Using Table I, the cardinality of the pre-image set $|A_{123}(\mathbf{b}, \mathbf{a}_3)| = 3^{t_{0,0}} 8^{t_{0,1}} 1^{t_{0,2}} 3^{t_{1,0}} 1^{t_{1,1}} 8^{t_{1,2}} 3^{t_{2,0}} = 3^{k'} 8^{t_{0,1}+t_{1,2}}$. Using this, we can find the value of the entropy as follows.

$$\begin{aligned}
& H_{|\mathcal{Z}|}(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|) \\
&= \sum_{u_0=0}^{t_{0,1}} \sum_{u_1=0}^{t_{1,2}} \binom{t_{0,1}}{u_0} 2^{u_0} \binom{t_{1,2}}{u_1} 2^{u_1} 3^{k'} \frac{4^{t_{0,1}+t_{1,2}}}{2^{u_0+u_1} 3^{k'} 8^{t_{0,1}+t_{1,2}}} \log_{|\mathcal{Z}|} \left(\frac{2^{u_0+u_1} 3^{k'} 8^{t_{0,1}+t_{1,2}}}{4^{t_{0,1}+t_{1,2}}} \right) \\
&= \frac{(t_{0,1} + t_{1,2}) \log_{|\mathcal{Z}|} 2 + k' \log_{|\mathcal{Z}|} 3}{2^{t_{0,1}+t_{1,2}}} \sum_{u_0=0}^{t_{0,1}} \sum_{u_1=0}^{t_{1,2}} \binom{t_{0,1}}{u_0} \binom{t_{1,2}}{u_1} + \frac{\log_{|\mathcal{Z}|} 2}{2^{t_{0,1}+t_{1,2}}} \sum_{u_0=0}^{t_{0,1}} \sum_{u_1=0}^{t_{1,2}} \binom{t_{0,1}}{u_0} u_0 \binom{t_{1,2}}{u_1} \\
&\quad + \frac{\log_{|\mathcal{Z}|} 2}{2^{t_{0,1}+t_{1,2}}} \sum_{u_0=0}^{t_{0,1}} \sum_{u_1=0}^{t_{1,2}} \binom{t_{0,1}}{u_0} u_1 \binom{t_{1,2}}{u_1} \\
&= k' \log_{|\mathcal{Z}|} 3 + 1.5(t_{0,1} + t_{1,2}) \log_{|\mathcal{Z}|} 2 \\
&= k \log_{|\mathcal{Z}|} 3 + (1.5 \log_{|\mathcal{Z}|} 2 - \log_{|\mathcal{Z}|} 3)(t_{0,1} + t_{1,2}) - (t_{0,2} + t_{1,1}) \log_{|\mathcal{Z}|} 3. \tag{10}
\end{aligned}$$

B. Value of α

Having evaluated $H(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|)$, we can find the value of α as used in equation (2) in the following manner.

$$\begin{aligned}
& H(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k), X_3^k) \\
&= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \sum_{\mathbf{x}_3} \Pr\{X_3^k = \mathbf{x}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} H(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}, X_3^k = \mathbf{x}_3) \\
&\geq \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \sum_{\mathbf{x}_3} \Pr\{X_3^k = \mathbf{x}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} H(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|) \triangleq \alpha k. \tag{11}
\end{aligned}$$

The value of $\Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\}$ can be found from Table I and the i.i.d. uniform assumption on the message tuples. The value of $\Pr\{X_3^k = \mathbf{x}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\}$ can be evaluated by finding the ratio of the cardinalities of two pre-image sets, i.e., $|A_{123}(\mathbf{b}, \mathbf{x}_3)| / |A_{123}(\mathbf{b})|$. We carry out this computation for the illustrated demand function below.

Illustration 6: Consider a realization \mathbf{b} with m_1 1's, m_2 2's and $k - m_1 - m_2$ 0's. Then we have that $\Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} = (12/27)^{m_1} (3/27)^{m_2} (12/27)^{k-m_1-m_2}$. The number of different demand function realizations which have the same number of 1's, 2's and 0's is $\binom{k}{m_1, m_2, k-m_1-m_2}$. Consider a realization $\mathbf{a}_3 \in A_3(\mathbf{b})$ of the message X_3^k . The number of message tuples that result in the demand function value \mathbf{b} and have their X_3^k realization as \mathbf{a}_3 is $|A_{123}(\mathbf{b}, \mathbf{a}_3)| = 3^{k'} 8^{t_{0,1}+t_{1,2}}$, as evaluated in the previous illustration. The number of message tuples in the pre-image set $A_{123}(\mathbf{b})$ can be evaluated using Table I as $|A_{123}(\mathbf{b})| = 12^{m_1} 3^{m_2} 12^{k-m_1-m_2} = 3^k 4^{k-m_2}$. Because of the uniform i.i.d. assumption, we have that

$$\begin{aligned}
\Pr\{X_3^k = \mathbf{a}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} &= \frac{|A_{123}(\mathbf{b}, \mathbf{a}_3)|}{|A_{123}(\mathbf{b})|} = \frac{3^{k'} 8^{t_{0,1}+t_{1,2}}}{3^k 4^{k-m_2}} \\
&= (1/4)^{k-m_1-m_2-t_{0,1}-t_{0,2}} (2/3)^{t_{0,1}} (1/12)^{t_{0,2}} (1/4)^{m_1-t_{1,1}-t_{1,2}} (2/3)^{t_{1,2}} (1/12)^{t_{1,1}}.
\end{aligned}$$

Using equations (11), (10) and the probabilities computed above, the value of α is found in Appendix C to be

$$\alpha = \frac{8}{9} \log_{|\mathcal{Z}|} 2 + \frac{4}{12} \log_{|\mathcal{Z}|} 3. \tag{12}$$

Using this value of α in equation (2), we get that

$$\frac{R_1 + R_2}{2} + \epsilon \geq \frac{\alpha + (\log_{|\mathcal{Z}|} 9 - \frac{8}{9} \log_{|\mathcal{Z}|} 4)}{2 \log_{|\mathcal{Z}|} |\mathcal{A}|} = \frac{\frac{8}{9} \log_{|\mathcal{A}|} 2 + \frac{1}{3} \log_{|\mathcal{A}|} 3 + 2 \log_{|\mathcal{A}|} 3 - \frac{16}{9} \log_{|\mathcal{A}|} 2}{2} \approx \frac{1.7725}{2}.$$

C. Example demand function: sum over $GF(2)$

Suppose the messages $X_1, X_2, X_3 \in \{0, 1\}$ and the terminal wants to compute their finite field sum over $GF(2)$. For this demand function, we demonstrate that the outer bound to the rate region is tight. We assume that the alphabet used for communication is $\mathcal{Z} = \{0, 1\}$. For X_3 we have that $0 \neq 1$, and thus for a function $g(X_3^k)$ that returns the equivalence class that X_3^k belongs to, we have that $H(g(X_3^k)) = k$. Thus from Lemma 2 we get that $R_{31} + R_{32} \geq 1$.

For X_u , $u = 1$ or 2 , using the values of the finite field sum, we obtain the following partitions

$$\stackrel{0}{=} |_u : \{0\} \cup \{1\} \quad \text{and} \quad \stackrel{1}{=} |_u : \{0\} \cup \{1\}.$$

Thus, for a $GF(2)$ -sum realization \mathbf{b} and $\mathbf{a}_3 \in A_3(\mathbf{b})$, the equivalence classes satisfy $|\text{Cl}_1^{(v)}(\mathbf{a}_3)| = |\text{Cl}_2^{(w)}(\mathbf{a}_3)| = 1$ for any class index v or w and $V_1(\mathbf{a}_3) = V_2(\mathbf{a}_3) = 2^k$. Thus the vector $\mathbf{d}_u(\mathbf{a}_3)$ defined in Lemma 5 is $\mathbf{1}_{2^k}$, and hence $H(\mathbf{d}_u(\mathbf{a}_3)/2^k) = k$, giving us the value of γ as

$$\gamma = 1 \implies R_u + \epsilon \geq 1.$$

For any value of X_1^k, X_3^k and $f(X_1^k, X_2^k, X_3^k)$ the value of X_2^k is fixed by $X_2^k = f(X_1^k, X_2^k, X_3^k) - X_1^k - X_3^k$, where the subtraction operations are also over $GF(2)$. Hence we have that

$$h_{\mathbf{a}_3}(\mathbf{v}, \mathbf{w}) = 1 \text{ for every } \mathbf{a}_3 \in A_3(\mathbf{b}) \text{ and every } (\text{Cl}_1^{(v)}, \text{Cl}_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b}).$$

We enumerate the different (X_1, X_2) pairs that result in $b^{(i)}$ for the realization $a_3^{(i)}$ in different cases as below.

- If $i \in \mathcal{I}_0(\mathbf{b}) \cap \mathcal{J}_0(\mathbf{a}_3)$: $(X_1, X_2) \in \{(0, 0), (1, 1)\}$.
- If $i \in \mathcal{I}_0(\mathbf{b}) \cap \mathcal{J}_1(\mathbf{a}_3)$: $(X_1, X_2) \in \{(0, 1), (1, 0)\}$.
- If $i \in \mathcal{I}_1(\mathbf{b}) \cap \mathcal{J}_0(\mathbf{a}_3)$: $(X_1, X_2) \in \{(0, 1), (1, 0)\}$.
- If $i \in \mathcal{I}_1(\mathbf{b}) \cap \mathcal{J}_1(\mathbf{a}_3)$: $(X_1, X_2) \in \{(0, 0), (1, 1)\}$.

Since there are two choices in each case, we have that $|A_{123}(\mathbf{b}, \mathbf{a}_3)| = 2^k$. Thus we have that $h_{\mathbf{b}, \mathbf{a}_3}/|A_{123}(\mathbf{b}, \mathbf{a}_3)| = \mathbf{1}_{2^k}/2^k$ which gives the value of α as

$$\alpha = 1 \implies (R_1 + R_2)/2 + \epsilon \geq (1 + H(f(X_1^k, X_2^k, X_3^k))/k)/2 = (1 + k/k)/2 = 1.$$

We describe simple network code that allows t to compute the $GF(2)$ -sum by carrying out the componentwise addition $\mathbf{Z}_1(X_1^k, \mathbf{Z}_{31}(X_3^k)) + \mathbf{Z}_2(X_2^k, \mathbf{Z}_{32}(X_3^k))$. The codewords used, for any $c \in \mathbb{Z}_{\geq 0}$ such that $c \leq k$, are

$$\begin{aligned} \mathbf{Z}_{31}(X_3^k) &= (X_3^{(1)}, X_3^{(2)}, \dots, X_3^{(c)}), \quad \mathbf{Z}_{32}(X_3^k) = (X_3^{(c+1)}, X_3^{(c+2)}, \dots, X_3^{(k)}), \\ \mathbf{Z}_1(X_1^k, \mathbf{Z}_{31}(X_3^k)) &= (X_1^{(1)} + X_3^{(1)}, \dots, X_1^{(c)} + X_3^{(c)}, X_1^{(c+1)}, \dots, X_1^{(k)}), \\ \mathbf{Z}_2(X_2^k, \mathbf{Z}_{32}(X_3^k)) &= (X_2^{(1)}, \dots, X_2^{(c)}, X_2^{(c+1)} + X_3^{(c+1)}, \dots, X_2^{(k)} + X_3^{(k)}). \end{aligned}$$

All operations above are over $GF(2)$. Then $\mathbb{E}\ell(\mathbf{Z}_{31}) = c, \mathbb{E}\ell(\mathbf{Z}_{32}) = k - c$. If $X_1, X_3 \sim \text{Unif}\{0, 1\}$ then $X_1 + X_3 \sim \text{Unif}\{0, 1\}$ and similarly for $X_2 + X_3$. Hence $\mathbb{E}\ell(\mathbf{Z}_1) = \mathbb{E}\ell(\mathbf{Z}_2) = k$. These imply that the rate tuple achieved is

$$(R_{31}, R_{32}, R_1, R_2) = (c/k, 1 - c/k, 1, 1),$$

which matches the lower bounds derived above.

D. Example demand function: arithmetic sum

Suppose the message alphabet is $\mathcal{A} = \{0, 1\}$, such that the messages X_1, X_2, X_3 are independent bits each equally likely to be 0 or 1. The demand function $f(X_1, X_2, X_3) = X_1 + X_2 + X_3$ is the sum of the messages over the integers, such that $\mathcal{B} = \{0, 1, 2, 3\}$. We use the codeword alphabet $\mathcal{Z} = \{0, 1\}$. This case of arithmetic sum computation in the variable-length network code framework was considered in [21] and we recover the results there in our general framework. For a given value of the arithmetic sum $\mathbf{b} \in \{0, 1, 2, 3\}^k$ the set $A_3(\mathbf{b})$ of valid realizations for X_3^k can be described as

$$A_3(\mathbf{b}) = \{\mathbf{a}_3 \in \{0, 1\}^k : a_3^{(i)} = 0 \text{ if } \mathbf{b}^{(i)} = 0 \text{ and } a_3^{(j)} = 1 \text{ if } \mathbf{b}^{(j)} = 3 \text{ for all } i, j \in \{1, 2, \dots, k\}\}.$$

We consider the equivalence relation $\equiv_{\mathbf{1}}^{\mathbf{a}_3}$ for the arithmetic sum demand function. Then

$$\mathbf{x}_1 \equiv_{\mathbf{1}}^{\mathbf{a}_3} \mathbf{y}_1 \iff \mathbf{x}_1 = \mathbf{y}_1,$$

because if $\mathbf{x}_1 \neq \mathbf{y}_1$ then for all $\mathbf{x}_2 \in \{0, 1\}^k$ we have that $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{a}_3 \neq \mathbf{y}_1 + \mathbf{x}_2 + \mathbf{a}_3$. A similar conclusion also holds true for the $\equiv_{\mathbf{2}}^{\mathbf{a}_3}$ relation. Thus for both $u = 1$ and 2 , we have that $|\text{Cl}_u^{(v)}(\mathbf{a}_3)| = 1$ for any class index v and $V_u(\mathbf{a}_3) = 2^k$. We use the alphabet $\mathcal{Z} = \{0, 1\}$ for communication. The vector $\mathbf{d}_u(\mathbf{a}_3)$ defined in Lemma 5 satisfies $\mathbf{d}_u(\mathbf{a}_3) = \mathbf{1}_{2^k}$ in this case, where $\mathbf{1}_t$ indicates a vector of ones with length t . Using this in equation (5), we obtain the value of γ as

$$\gamma = \frac{1}{k} \sum_{\mathbf{a}_3 \in \mathcal{A}^k} \Pr\{X_3^k = \mathbf{a}_3\} H_{|\mathcal{Z}|} \left(\frac{\mathbf{d}_u(\mathbf{a}_3)}{|\mathcal{A}|^k} \right) = \frac{1}{k} \cdot k = 1,$$

and thus $R_u + \epsilon \geq 1/\log_{|\mathcal{Z}|} |\mathcal{A}| = 1$.

For any value of X_1^k, X_3^k and $f(X_1^k, X_2^k, X_3^k)$ the value of X_2^k is fixed by $X_2^k = f(X_1^k, X_2^k, X_3^k) - X_1^k - X_3^k$. Hence, for every $(\text{Cl}_1^{(v)}, \text{Cl}_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b})$ there is exactly one message tuple whose X_1^k and X_2^k belong to the equivalence classes $\text{Cl}_1^{(v)}(\mathbf{a}_3)$ and $\text{Cl}_2^{(w)}(\mathbf{a}_3)$ respectively. Thus we have that

$$h_{\mathbf{a}_3}(\mathbf{v}, \mathbf{w}) = 1 \text{ for every } \mathbf{a}_3 \in A_3(\mathbf{b}) \text{ and every } (\text{Cl}_1^{(v)}, \text{Cl}_2^{(w)}) \in \mathcal{V}_{12}(\mathbf{a}_3, \mathbf{b}).$$

Consider an arithmetic sum realization \mathbf{b} with m_0 0's, m_1 1's, m_2 2's and $k - m_0 - m_1 - m_2$ 3's. Define $t_{p,q} \triangleq |\mathcal{I}_p(\mathbf{b}) \cap \mathcal{J}_q(\mathbf{a}_3)|$ for every $p \in \{0, 1, 2, 3\}$ and $q \in \{0, 1\}$. Then for any choice of \mathbf{b} and $\mathbf{a}_3 \in A_3(\mathbf{b})$, $t_{0,1} = t_{3,0} = 0$. The cardinality of the pre-image set $|A_{123}(\mathbf{b}, \mathbf{a}_3)| = 2^{t_{1,0} + t_{2,1}} = 2^{t_{1,0} + m_2 - t_{2,0}}$. The value of the entropy $H(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|) = t_{1,0} + m_2 - t_{2,0}$. From the function definition, we can check that $|A_{123}(\mathbf{b})| = 3^{m_1 + m_2}$. Thus we have that

$$\Pr\{X_3^k = \mathbf{x}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} = \frac{|A_{123}(\mathbf{b}, \mathbf{a}_3)|}{|A_{123}(\mathbf{b})|} = \frac{2^{t_{1,0} + m_2 - t_{2,0}}}{3^{m_1 + m_2}}.$$

Then the value of α can be found as follows.

$$\begin{aligned} \alpha k &= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \sum_{\mathbf{a}_3 \in A_3(\mathbf{b})} \Pr\{X_3^k = \mathbf{a}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} H(\mathbf{h}_{\mathbf{b}, \mathbf{a}_3} / |A_{123}(\mathbf{b}, \mathbf{a}_3)|) \\ &= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \sum_{t_{1,0}=0}^{m_1} \sum_{t_{2,0}=0}^{m_2} \frac{m_1! (2/3)^{t_{1,0}} (1/3)^{m_1 - t_{1,0}}}{t_{1,0}! (m_1 - t_{1,0})!} \frac{m_2! (2/3)^{m_2 - t_{2,0}} (1/3)^{t_{2,0}}}{t_{2,0}! (m_2 - t_{2,0})!} (t_{1,0} + m_2 - t_{2,0}) \\ &= \sum_{m_1=0}^k \sum_{m_2=0}^{k-m_1} \frac{k! 2^{k-m_1-m_2}}{m_1! m_2! (k-m_1-m_2)!} \left(\frac{1}{8}\right)^{k-m_1-m_2} \left(\frac{3}{8}\right)^{m_1+m_2} \frac{2(m_1+m_2)}{3} \\ &= \frac{2/3}{4^k} \sum_{m_1=0}^k \sum_{m_2=0}^{k-m_1} \frac{k! (m_1+m_2)}{m_1! m_2! (k-m_1-m_2)!} \left(\frac{3}{2}\right)^{m_1+m_2} = \frac{2/3}{4^k} 2 \cdot \frac{3}{2} k 4^{k-1} = 0.5k. \end{aligned}$$

Putting this value of α in equation (2), we get that

$$\frac{R_1 + R_2}{2} + \epsilon \geq \frac{0.5 + 3 - 0.75 \log 3}{2} \approx \frac{2.31128}{2}.$$

Remark 1: We note that the lower bound for the sum rate shown above is tighter than the bound $R_1 + R_2 > 2.25$ obtained in [21] for the same problem. The inequalities considered there were similar to those used in arriving at the sum rate lower bound in equation (2), however, in [21] they did not include X_3^k in the conditioning while lower bounding $H(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k))$ as done here. Instead, they directly lower bounded $H(\mathbf{Z}_1, \mathbf{Z}_2 | f(X_1^k, X_2^k, X_3^k))$ using a so-called *clumpy* distribution, which is a p.m.f. that majorizes any valid conditional p.m.f. of the pair $(\mathbf{Z}_1, \mathbf{Z}_2)$ given $f(X_1^k, X_2^k, X_3^k)$. Since only a lower bound to the entropy of the clumpy distribution was obtained in [21], the corresponding lower bound for the sum rate ends up being looser than the value we obtain here.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have described a procedure to obtain an outer bound for the rate region (in the setup of [22]) for computing a function with zero-error over a simple DAG network. The demand function can be an arbitrary discrete-valued function and only needs to be specified as a function table. For computing the arithmetic sum, we show that the outer bound obtained is tighter than the one in [21]. For computing the $GF(2)$ -sum, we show that the lower bounds for the rate tuple obtained using our procedure can also be achieved by a simple network code. Our method uses the equivalence relations defined in [19] as adapted to the specific DAG network considered here. Assuming a independent, uniform probability distribution for each of the messages, we compute the probability that the messages belong to a particular equivalence class. These are used in obtaining a lower bound to the conditional entropy of the descriptions transmitted on the edges, which imply an outer bound to the rate region.

There are several opportunities for future work. Our investigation reveals that the analysis of even very simple, non-tree networks require the usage of fairly nontrivial techniques. Nevertheless, it is evident that taking into account the compressibility of the sources allows for higher computation rates. It would be interesting to consider the application of our techniques (or simplifications thereof) to more complicated topologies. It may also be fruitful to consider specific functions of interest for which the equivalence classes (used extensively in Section III) have a simpler characterization.

REFERENCES

- [1] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 755-764, April 2005.
- [2] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, Jan 2012.
- [3] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Trans. on Info. Th.*, vol. 25, no. 2, pp. 219-221, Mar 1979.
- [4] H. Witsenhausen, "The zero-error side information problem and chromatic numbers (Corresp.)," *IEEE Trans. on Info. Th.*, vol. 22, no. 5, pp. 592-593, September 1976.
- [5] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. on Info. Th.*, vol. 42, no. 5, pp. 1329-1339, Sep 1996.
- [6] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. on Info. Th.*, vol. 44, no. 6, pp. 2207-2229, Oct 1998.
- [7] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. on Info. Th.*, vol. 47, no. 3, pp. 903-917, Mar 2001.

- [8] V. Doshi, D. Shah, M. Médard and M. Effros, “Functional Compression Through Graph Coloring,” *IEEE Trans. on Info. Th.*, vol. 56, no. 8, pp. 3901-3917, Aug. 2010.
- [9] T. S. Han and K. Kobayashi, “A dichotomy of functions $F(X, Y)$ of correlated sources (X, Y) from the viewpoint of the achievable region,” *IEEE Trans. on Info. Th.*, vol. 33, no. 1, pp. 69-76, January 1987.
- [10] S. Feizi and M. Médard, “On Network Functional Compression,” *IEEE Trans. on Info. Th.*, vol. 60, no. 9, pp. 5387-5401, Sept. 2014.
- [11] M. Sefidgaran and A. Tchamkerten, “Distributed Function Computation Over a Rooted Directed Tree,” *IEEE Trans. on Info. Th.*, vol. 62, no. 12, pp. 7135-7152, Dec. 2016.
- [12] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, “Network information flow,” *IEEE Trans. on Info. Th.*, vol. 46, no. 4, pp. 1204-1216, Jul 2000.
- [13] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Trans. on Networking*, vol. 11, no. 5, pp. 782-795, Oct. 2003.
- [14] H. Kowshik and P. R. Kumar, “Optimal Function Computation in Directed and Undirected Graphs,” in *IEEE Trans. on Info. Th.*, vol. 58, no. 6, pp. 3407-3418, June 2012.
- [15] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, “Network coding for computing: Cut-set bounds,” *IEEE Trans. on Info. Th.*, vol. 57, no. 2, pp. 1015-1030, Feb 2011.
- [16] J. Cannons, R. Dougherty, C. Freiling and K. Zeger, “Network routing capacity,” *IEEE Trans. on Info. Th.*, vol. 52, no. 3, pp. 777-788, March 2006.
- [17] R. Dougherty, C. Freiling and K. Zeger, “Unachievability of network coding capacity,” *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2365-2372, June 2006.
- [18] C. Huang, Z. Tan, and S. Yang, “Upper bound on function computation in directed acyclic networks,” in *IEEE Info. Th. Workshop*, April 2015, pp. 1-5.
- [19] X. Guang, S. Yang, and C. Li, “An improved upper bound on network function computation using cut-set partition,” in *IEEE Info. Th. Workshop*, Sept 2016, pp. 11-15.
- [20] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, ser. Springer Series in Statistics. Springer New York, 2011.
- [21] A. Tripathy and A. Ramamoorthy, “On computation rates for arithmetic sum,” in *IEEE Intl. Symposium on Info. Th.*, July 2016, pp. 2354-2358.
- [22] L. Song, R. W. Yeung and N. Cai, “Zero-error network coding for acyclic networks,” in *IEEE Trans. on Info. Th.*, vol. 49, no. 12, pp. 3129-3139, Dec. 2003.
- [23] T. Cover and J. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.
- [24] S. Leung-Yan-Cheong and T. Cover, “Some equivalences between Shannon entropy and Kolmogorov complexity,” in *IEEE Trans. on Info. Th.*, vol. 24, no. 3, pp. 331-338, May 1978.
- [25] X. Guang, R. W. Yeung, S. Yang and C. Li, “Improved upper bound on the network function computing capacity,” online at <http://arxiv.org/abs/1710.02252>.
- [26] R. W. Yeung, *A First Course in Information Theory*, ser. Information Technology: Transmission, Processing and Storage. Springer US, 2002.
- [27] A. Ramamoorthy and M. Langberg, “Communicating the sum of sources over a network,” *IEEE J. Select. Areas Comm.*, vol. 31, no. 4, pp. 655-665, 2013.
- [28] B. K. Rai and B. K. Dey, “On network coding for sum-networks,” *IEEE Trans. on Info. Th.*, vol. 58, no. 1, pp. 50 -63, 2012.
- [29] A. Tripathy and A. Ramamoorthy, “Sum-networks from incidence structures: construction and capacity analysis,” *IEEE Trans. on Info. Th.*, vol. 64, no. 5, pp. 3461 -3480, 2018.
- [30] R. Dougherty, C. Freiling, and K. Zeger, “Insufficiency of linear coding in network information flow,” *IEEE Trans. on Info. Th.*, vol. 51, no. 8, pp. 2745-2759, Aug. 2005.
- [31] S. Huang and A. Ramamoorthy, “An achievable region for the double unicast problem based on a minimum cut analysis,” *IEEE Trans. on Comm.*, vol. 61, no. 7, pp. 2890-2899, 2013.
- [32] —, “On the multiple unicast capacity of 3-source, 3-terminal directed acyclic networks,” *IEEE/ACM Trans. on Networking*, vol. 22, no.1, pp. 285-299, 2014.

APPENDIX A PROOF OF LEMMA 1

Proof: Let $l_1 \leq l_2 \leq \dots \leq l_{\max}$ be the lengths over \mathcal{Z} of the best non-singular code for \mathcal{Z} . We demonstrate a function g such that the set of lengths $\{g(l_i) : i = 1, 2, \dots\}$ satisfy Kraft’s inequality, i.e., $\sum_{i=1}^{\infty} |\mathcal{Z}|^{-g(l_i)} \leq 1$. Choose $g(l_i) \triangleq l_i + 2\lfloor \log_{|\mathcal{Z}|}(l_i + |\mathcal{Z}| - 1) \rfloor$. Since there are $|\mathcal{Z}|^{l_i}$ different non-singular codewords with length l_i and if $l_{\max} > l_i$ then all of them must have been used in the best non-singular code, we have that

$$\begin{aligned} \sum_i |\mathcal{Z}|^{-l_i - 2\lfloor \log_{|\mathcal{Z}|}(l_i + |\mathcal{Z}| - 1) \rfloor} &= \sum_{l_1}^{l_{\max}} |\mathcal{Z}|^{l_i} |\mathcal{Z}|^{-l_i - 2\lfloor \log_{|\mathcal{Z}|}(l_i + |\mathcal{Z}| - 1) \rfloor} \leq \sum_{l=1}^{\infty} |\mathcal{Z}|^{-(2\lfloor \log_{|\mathcal{Z}|}(l + |\mathcal{Z}| - 1) \rfloor)} \\ &= \sum_{l=1}^{|\mathcal{Z}|^2 - |\mathcal{Z}|} \frac{1}{|\mathcal{Z}|^2} + \sum_{l=|\mathcal{Z}|^2 - |\mathcal{Z}| + 1}^{|\mathcal{Z}|^3 - |\mathcal{Z}|} \frac{1}{|\mathcal{Z}|^4} + \sum_{l=|\mathcal{Z}|^3 - |\mathcal{Z}| + 1}^{|\mathcal{Z}|^4 - |\mathcal{Z}|} \frac{1}{|\mathcal{Z}|^6} + \dots \\ &= \frac{|\mathcal{Z}| - 1}{|\mathcal{Z}|} + \frac{|\mathcal{Z}| - 1}{|\mathcal{Z}|^2} + \frac{|\mathcal{Z}| - 1}{|\mathcal{Z}|^3} + \dots = 1. \end{aligned}$$

The above is also true if $l_{\max} \rightarrow \infty$. Thus there exists a uniquely decodable code for \mathcal{Z} whose codeword lengths are $\{\lceil g(l_i) \rceil : i = 1, 2, \dots\}$. Using random variable ℓ to denote the lengths of the codewords in the best non-singular code for \mathcal{Z} , we have that

$$\begin{aligned} H_{|\mathcal{Z}|}(\mathcal{Z}) &\leq \mathbb{E}(\ell + 2\lfloor \log_{|\mathcal{Z}|}(\ell + |\mathcal{Z}| - 1) \rfloor) \leq \mathbb{E} \ell + 2 \mathbb{E} \log_{|\mathcal{Z}|}(\ell + |\mathcal{Z}| - 1) \stackrel{(i)}{\leq} \mathbb{E} \ell + 2 \log_{|\mathcal{Z}|}(\mathbb{E} \ell + |\mathcal{Z}| - 1) \\ &\leq \mathbb{E} \ell + 2 \log_{|\mathcal{Z}|}(1 + H_{|\mathcal{Z}|} + |\mathcal{Z}| - 1) \implies \mathbb{E} \ell \geq H_{|\mathcal{Z}|}(\mathcal{Z}) - 2 \log_{|\mathcal{Z}|}(|\mathcal{Z}| + H_{|\mathcal{Z}|}(\mathcal{Z})), \end{aligned}$$

where inequality (i) above is true due to Jensen’s inequality. ■

APPENDIX B
PROOF OF LEMMA 4

Proof: We use the same argument for $u = 1$ and 2 . Partition the set \mathcal{A}^k into $\prod_{i=1}^k V_u(a_3^{(i)})$ disjoint subsets based on the equivalence relation $\stackrel{a_3^{(i)}}{\equiv} |_{u}$ in each component $i \in \{1, 2, \dots, k\}$. Every element $(x_u^{(1)}, x_u^{(2)}, \dots, x_u^{(k)}) \in \mathcal{A}^k$ belongs to a subset in the partition. Suppose the number of distinct \mathcal{Z}_u -labels transmitted on (s_u, t) is strictly less than $\prod_{i=1}^k V_u(a_3^{(i)})$. Then by the pigeon-hole principle, there exist two elements $\mathbf{x}_u \triangleq (x_u^{(1)}, x_u^{(2)}, \dots, x_u^{(k)})$ and $\mathbf{y}_u \triangleq (y_u^{(1)}, y_u^{(2)}, \dots, y_u^{(k)})$ that belong to different equivalence relation subsets of \mathcal{A}^k but are given the same \mathcal{Z}_u -label. Let $J \subseteq \{1, 2, \dots, k\}$ be the index set collecting all indices j such that $x_u^{(j)} \stackrel{a_3^{(j)}}{\not\equiv} y_u^{(j)} |_{u}$. Since \mathbf{x}_u and \mathbf{y}_u belong to different equivalence relation partitions, $J \neq \emptyset$ and for every $j \in J$ there exists, by definition, a $a_v^{(j)} \in \mathcal{A}, v \in \{1, 2\} \setminus u$ such that

$$f\left(x_u^{(j)}, a_v^{(j)}, a_3^{(j)}\right) \neq f\left(y_u^{(j)}, a_v^{(j)}, a_3^{(j)}\right).$$

Then consider the following two different scenarios of k independent message realizations.

- (I) $X_u^k = \mathbf{x}_u, X_v^k = \mathbf{x}_v, X_3^k = \mathbf{a}_3$
 (II) $X_u^k = \mathbf{y}_u, X_v^k = \mathbf{x}_v, X_3^k = \mathbf{a}_3$,

where \mathbf{x}_v is such that $x_v^{(j)} = a_v^{(j)}$ for every $j \in J$. Note that the realizations of X_v^k and X_3^k are the same in both cases, and hence so is the label transmitted on the (s_v, t) edge. On the other hand, by assumption we have that the label transmitted on the edge (s_u, t) is the same in both cases as well. Then the terminal cannot recover the correct value of the demand function for the components in the set J , as the $(\mathcal{Z}_1, \mathcal{Z}_2)$ -labels received are the same but the function values at the components in J are different by choice of $\mathbf{x}_u, \mathbf{y}_u$. This contradicts the fact that the network code allows t to recover $f(X_1^k, X_2^k, X_3^k)$ with zero error. The two scenarios considered above have $\mathbf{x}_u \stackrel{a_3}{\not\equiv} \mathbf{y}_u |_{u}$ and thus also give a proof for the second statement of the lemma. \blacksquare

APPENDIX C
CALCULATION FOR EQUATION (12)

Using equations (11), (10) and the probabilities computed in illustration 6, we can find the value of α as follows.

$$\begin{aligned} & \alpha k \\ &= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \sum_{\mathbf{x}_3 \in \mathcal{A}_3(\mathbf{b})} \Pr\{X_3^k = \mathbf{x}_3 | f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \left(k \log_{|\mathcal{Z}|} 3 - (t_{0,2} + t_{1,1}) \log_{|\mathcal{Z}|} 3 \right. \\ & \qquad \qquad \qquad \left. + (1.5 \log_{|\mathcal{Z}|} 2 - \log_{|\mathcal{Z}|} 3)(t_{0,1} + t_{1,2}) \right) \\ &= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \left[k \log_{|\mathcal{Z}|} 3 \right. \\ & \quad + \sum_{t_{1,1}=0}^{m_1} \sum_{t_{1,2}=0}^{m_1-t_{1,1}} \binom{m_1}{t_{1,1}, t_{1,2}, m_1-t_{1,1}-t_{1,2}} \left(\frac{1}{4}\right)^{m_1-t_{1,1}-t_{1,2}} \left(\frac{2}{3}\right)^{t_{1,2}} \left(\frac{1}{12}\right)^{t_{1,1}} \left((\log_{|\mathcal{Z}|} \frac{2^{1.5}}{3}) t_{1,2} - t_{1,1} \log_{|\mathcal{Z}|} 3 \right) \\ & \quad \cdot \sum_{t_{0,1}=0}^{k-m_1-m_2} \sum_{t_{0,2}=0}^{k-m_1-m_2-t_{0,1}} \binom{k-m_1-m_2}{t_{0,1}, t_{0,2}, k-m_1-m_2-t_{0,1}-t_{0,2}} \left(\frac{1}{4}\right)^{k-m_1-m_2-t_{0,1}-t_{0,2}} \left(\frac{2}{3}\right)^{t_{0,1}} \left(\frac{1}{12}\right)^{t_{0,2}} \\ & \quad + \sum_{t_{0,1}=0}^{k-m_1-m_2} \sum_{t_{0,2}=0}^{k-m_1-m_2-t_{0,1}} \binom{k-m_1-m_2}{t_{0,1}, t_{0,2}, m_1-t_{0,1}-t_{0,2}} \left(\frac{1}{4}\right)^{k-m_1-m_2-t_{0,1}-t_{0,2}} \left(\frac{2}{3}\right)^{t_{0,1}} \left(\frac{1}{12}\right)^{t_{0,2}} \\ & \qquad \qquad \qquad \cdot \left((\log_{|\mathcal{Z}|} \frac{2^{1.5}}{3}) t_{0,1} - t_{0,2} \log_{|\mathcal{Z}|} 3 \right) \\ & \quad \cdot \sum_{t_{1,1}=0}^{m_1} \sum_{t_{1,2}=0}^{m_1-t_{1,1}} \binom{m_1}{t_{1,1}, t_{1,2}, m_1-t_{1,1}-t_{1,2}} \left(\frac{1}{4}\right)^{m_1-t_{1,1}-t_{1,2}} \left(\frac{2}{3}\right)^{t_{1,2}} \left(\frac{1}{12}\right)^{t_{1,1}} \left. \right] \\ &= \sum_{\mathbf{b}} \Pr\{f(X_1^k, X_2^k, X_3^k) = \mathbf{b}\} \left[k \left(\log_{|\mathcal{Z}|} 2 + \frac{\log_{|\mathcal{Z}|} 3}{4} \right) + m_2 \left(-\log_{|\mathcal{Z}|} 2 + \frac{9}{12} \log 3 \right) \right] \\ &= \sum_{m_1=0}^k \sum_{m_2=0}^{k-m_1} \binom{k}{m_1, m_2, k-m_1-m_2} \left(\frac{4}{9}\right)^{k-m_1-m_2} \left(\frac{4}{9}\right)^{m_1} \left(\frac{1}{9}\right)^{m_2} \left[k \left(\log_{|\mathcal{Z}|} 2 + \frac{\log_{|\mathcal{Z}|} 3}{4} \right) \right. \end{aligned}$$

$$\begin{aligned}
& + m_2 \left(-\log_{|z|} 2 + \frac{9}{12} \log_{|z|} 3 \right) \Big] \\
& = k \left(\log_{|z|} 2 + \frac{\log_{|z|} 3}{4} \right) \\
& \quad + \left(-\log_{|z|} 2 + \frac{9}{12} \log_{|z|} 3 \right) \sum_{m_1=0}^k \sum_{m_2=0}^{k-m_1} \binom{k}{m_1, m_2, k-m_1-m_2} m_2 \left(\frac{1}{9}\right)^{m_2} \left(\frac{4}{9}\right)^{k-m_2} \\
& = k \left(\log_{|z|} 2 + \frac{\log_{|z|} 3}{4} \right) + \left(-\log_{|z|} 2 + \frac{9}{12} \log_{|z|} 3 \right) \frac{k}{9}.
\end{aligned}$$