Linear Network Codes: A Unified Framework for Source, Channel, and Network Coding

Minkyu Kim

6.454 Fall 2003

1 Introduction

This report is a summary of the issue of separation and code design for network data transmission environments, as discussed in the paper by Effros *et al.* [1].

It is often considered that the failure of source-channel separation in network environments is a crucial obstacle in applying information theoretic tools in networks. However, with a consistent framework for digital networks, the source-channel separation turns out to be more robust than existing counterexamples may suggest. More interestingly, the separation of source and channel code design does not necessarily simplify the design of communication systems for digital networks; rather such decomposition of a problem into modular tasks may increase complexity as the decomposition imposes unnecessary conditions.

The network coding literature assumes independent data bits and lossless links, hence endorses a philosophy where source and channel coding are separated from network coding or routing. However, there are simple examples showing neither separate source-network coding nor separate channel-network coding techniques guarantee optimal performance. We further argue that the major challenge in design of network systems is the lack of separation of large networks into canonical elements such as simple multiple access or broadcast networks, rather than the lack of source-channel separation in networks.

2 Preliminaries

The network model to be considered here requires the same finite alphabet at all nodes and additionally allows erasures assumed to be channel-imposed, irreversible, and independent of the channel input. We deal with two important types of networks: multiple access and broadcast networks, and we argue that random linear codes are asymptotically optimal for those setups. This approach may be viewed, in the simplest way, as a generalization of information theoretic results known for single-receiver source codes and for single-transmitter, single-receiver channel codes. Also, from the networking perspective, it has an interpretation that compression, channel coding, and routing are not separable functions.

All results in this research are in their simplest forms. In particular, in all cases the source and channel alphabets are binary, including the erasure symbol if needed, and all results are stated for iid random processes. Also, all code constructions combine, for simplicity, random linear encoding with typical set decoding. However, all of the results may generalize widely from the forms stated here.

3 Single-Transmitter, Single-Receiver Networks

Given a single-transmitter, single-receiver network, source coding can be viewed as an extension of network coding to application with statistically dependent input symbols. A network code is said to accomplish optimal source coding on a noise-free network if that code can be used to transmit any source with entropy lower that the network capacity with asymptotically negligible error probability.

Let us begin by showing that the expected error probability of a randomly chosen linear source code with rate R tends to zero as n grows without bound for any source U with $H(U) < R$. The fixed-rate, linear encoder is independent of the source distribution, and we use distribution-dependent typical set decoders for simplicity.

Let a_n be an $\lceil nR \rceil \times n$ matrix with coefficients in the binary field $\mathbb{F}_2$. The encoder for the linear source code based on a_n is

$$\alpha_n(u^n) = a_n \mathbf{u},$$

where $u^n = \mathbf{u}^t \in (\mathbb{F}_2)^n$ is an arbitrary source sequence with blocklength n. The corresponding decoder is

$$\beta_n(v^{\lceil nR \rceil}) = \begin{cases} u^n & \text{if } u_n \in A_\epsilon^{(n)} \text{ and } a_n \mathbf{u} = \mathbf{v} \text{ and } \nexists \hat{u}^n \in A_\epsilon^{(n)} \cap \{u^n\} \text{ s.t. } a_n \hat{\mathbf{u}} = \mathbf{v} \\ \hat{U}^n & \text{otherwise,} \end{cases}$$

where $v^{\lceil nR \rceil}) = \mathbf{v}^t \in \mathbb{F}_2^{\lceil nR \rceil}$ and decoding to $\hat{U}^n$ denotes a random decoder output (which yields a decoding error by assumption). The error probability for source code a_n is

$$P_e(a_n) = \Pr(\beta_n(\alpha_n(U^n)) \neq U^n).$$

Then, we have the following source coding theorem:

Theorem 1 *Let U_1, U_2, ..., U_n be drawn iid according to distribution $p(u)$. Let $\{A_n\}_{n=1}^\infty$ be a sequence of rate-R linear source codes. Each A_n is an $\lceil nR \rceil \times n$ matrix with coefficients drawn iid Bernoulli(1/2). For any $R > H(U)$, $E[P_e(A_n)] \to 0$ as $n \to 0$.*

While Theorem 1 shows that linear source codes are asymptotically optimal, it can be shown that any fixed linear code yields statistically dependent output symbols. Therefore, linear source codes cannot achieve the entropy bound for non-uniform sources since achieving the entropy bound would necessarily yield an incompressible data sequence.

Similarly to the case of source coding, channel coding also can be viewed as an extension of network coding to unreliable channels. Prior network coding results treat the issue of robust communication against non-ergodic link failures, but here we investigate ergodic failures. We say that a network code accomplished optimal channel coding on the given channel if the network code can be used to transmit any source with rate lower than the noisy channel capacity with asymptotically negligible error probability.

For linear channel coding for the erasure channel, we use an $n \times \lfloor nR \rfloor$ linear generator matrix b_n and a conceptually simple non-linear decoder. The linear channel encoder is defined by

$$\gamma(v^{\lfloor nR \rfloor}) = b_n \mathbf{v}.$$

For any channel output $y_n = \mathbf{y}^t \in \{0, 1, E\}^n$, define the decoder as

$$\delta_n(y^n) = \begin{cases} v^n & \text{if } (b_n \mathbf{v})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \text{ and } \nexists \hat{\mathbf{v}} \neq \mathbf{v} \text{ s.t. } (b_n \hat{\mathbf{v}})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \\ \hat{V}^{\lfloor n \rfloor} & \text{otherwise,} \end{cases}$$

where for any $\mathbf{v} \in \mathbb{F}_2^{\lfloor nR \rfloor}$, $(b_n\mathbf{v})_i$ is the ith component of the vector $b_n\mathbf{v}$. Again, decoding to $\hat{V}^{\lfloor n \rfloor}$ denotes a random decoder output. Then we have the following channel coding theorem:

Theorem 2 *Consider an erasure channel with input and output alphabets $\mathbb{F}_2$ and $\{0, 1, E\}$, respectively. The erasure sequence Z_1, Z_2, ... is drawn iid according to distribution $q(z)$, where $Z_i = 1$ denotes the erasure event, and $Z_i = 0$ designates a successful transmission. The channel noise is assumed independent of the channel input. Let $\{B_n\}_{n=1}^{\infty}$ describe a sequence of channel codes. Each B_n is an $n \times \lfloor nR \rfloor$ matrix with elements chosen iid Bernoulli(1/2). If $R < 1 - q(1)$, then $E[P_e(B_n)] \to 0$ as $n \to 0$.*

Given source code a_n and channel code b_n, the joint source-channel encoder multiplies the source input by a single $n \times n$ matrix $c_n = b_n a_n$ and transmits the output across the channel. The corresponding decoder is $\beta_n(\delta(\cdot))$. As an alternative to this concatenating method, we can design a joint source-channel code at random and decode in a single typical set decoding argument.

The joint source-channel code's encoder is defined as

$$\zeta(u^n) = c_n\mathbf{u},$$

for any channel output $y^n = \mathbf{y}^t \in \{0, 1, E\}$ the decoder is defined by

$$\eta_n(y^n) = \begin{cases} u^n & \text{if } (c_n\mathbf{u})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \text{ and } \nexists \hat{\mathbf{u}} \neq \mathbf{u} \text{ s.t. } (b_n\hat{\mathbf{u}})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \\ \hat{U}^n & \text{otherwise,} \end{cases}$$

Theorem 3 *Consider the random source U_1, U_2, ... drawn iid according to distribution $p(u)$, and let Z_1, Z_2, ... be the channel's random erasures, where Z_1, Z_2, ... are drawn iid according to distribution $q(z)$ and independent of the source. Let $\{C_n\}_{n=1}^{\infty}$ describe a sequence of joint source-channel codes. Each C_n is an $n \times n$ matrix with elements chosen iid Bernoulli(1/2). If $H(U) < 1 - q(1)$, then $E[P_e(C_n)] \to 0$ as $n \to 0$.*

We note that both the channel coding and the joint source-channel coding theorem extend easily to additive noise channels. For channel coding, let a_n be an $\lceil n(1-R) \rceil \times n$ matrix with coefficients in $\mathbb{F}_2$, which can be interpreted as a source code on the noise as will be shown. For any matrix a_n, we can design an $n \times \lfloor nR \rfloor$ matrix b_n such that b_n has full rank and $a_n b_n = \mathbf{0}$. Matrix b_n plays the role of the generator matrix for the desired channel code.

The channel encoder is defined by

$$\gamma(v^{n-k}) = b_n\mathbf{v},$$

and the channel output is

$$\mathbf{Y} = b_n\mathbf{v} + \mathbf{Z}.$$

In decoding the channel output, the receiver first multiplies $\mathbf{Y}$ by a_n to give

$$a_n\mathbf{Y} = a_n(b_n\mathbf{v} + \mathbf{Z}) = a_n\mathbf{Z}.$$

This, the decoding procedure involves applying source decoder β_n to $a_n\mathbf{Z}$. After subtracting the error estimate from $\mathbf{Y}$, the receiver can recover $\mathbf{v}$ from $b_n\mathbf{v}$ since b_n has full rank.

Theorem 4 *Consider an additive noise channel with input, output, and noise alphabets all equal to the binary field $\mathbb{F}_2$. Let noise Z_1, Z_2, ... be drawn iid according to distribution $q(z)$ and be independent of the channel input. Let $\{(B_n, A_n)\}_{n=1}^{\infty}$ describe a sequence of channel codes. Each A_n is $\lceil n(1-R) \rceil$ matrix with elements chosen iid Bernoulli(1/2), and each B_n is designed to match the corresponding A_n as above. If $R < 1 - H(Z)$, then $E[P_e(B_n, A_n)] \to 0$ as $n \to 0$.*

As before, we can consider a linear joint source-channel code for the additive noise channel as well. An encoder is defined by

$$\zeta(u^n) = c_n \mathbf{u},$$

and for given channel input $c_n\mathbf{u}$, the channel output is

$$\mathbf{Y} = c_n\mathbf{u} + \mathbf{Z}.$$

We define the decoder as

$$\eta_n(y^n) = \begin{cases} u^n & \text{if } u_n \in A_\epsilon^{(n)}(p) \text{ and } \exists z^n \in A_\epsilon^{(n)}(q) \text{ s.t. } c_n\mathbf{u} + \mathbf{z} = \mathbf{y} \\ & \quad \text{and } \nexists(\hat{\mathbf{u}}, \hat{\mathbf{z}}) \in (A_\epsilon^{(n)}(p) \cap \{\mathbf{u}\}^c) \times A_\epsilon^{(n)}(q) \text{ s.t. } c_n\hat{\mathbf{u}} + \hat{\mathbf{z}} = \mathbf{y} \\ \hat{U}^n & \text{otherwise.} \end{cases}$$

Theorem 5 *Consider the random source U_1, U_2, ... drawn iid according to distribution $p(u)$, and let Z_1, Z_2, ... be the channel's random additive noise, where Z_1, Z_2, ... are drawn iid according to distribution $q(z)$ and independent of the source. Let $\{C_n\}_{n=1}^\infty$ describe a sequence of joint source-channel codes. Each C_n is an $n \times n$ matrix with elements chosen iid Bernoulli(1/2). If $H(U) < 1 - H(Z)$, then $E[P_e(C_n)] \to 0$ as $n \to 0$.*

4 Multiple Access Systems

Given the results in the previous section, the same strategies can also be applied to multiple access systems. Let us first consider source coding. Given $\lceil nR_1 \rceil \times n$ matrix $a_{1,n}$ and $\lceil nR_2 \rceil \times n$ matrix $a_{2,n}$, we associate with $(a_{1,n}, a_{2,n})$ a blocklength-n, two transmitter, linear multiple access source code similarly as in the previous section except that now we have a tuple of codewords. Then we have the following source coding theorem:

Theorem 6 *Consider source sequence $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2})$, ... drawn iid according to distribution $p(u_1, u_2)$ on $(\mathbb{F}_2)^2$. Let $\{(A_{1,n}, A_{2,n})\}_{n=1}^\infty$ be a sequence of rate-(R_1, R_2) linear multiple-access source codes with coefficients coefficients drawn iid Bernoulli(1/2). For any rates*

$$\begin{aligned} R_1 &> H(U_1|U_2) \\ R_2 &> H(U_2|U_1) \\ R_1 + R_2 &> H(U_1, U_2), \end{aligned}$$

$E[P_e(A_{1,n}, A_{2,n})] \to 0$ as $n \to 0$.

For channel coding, we consider two additive multiple access channels shown in Figure 1. The first is the additive multiple access channel with erasures, and the second is with additive noise. Let X_1^n and X_2^n be the random channel inputs and Y^n be the corresponding channel output, then for the erasure channel model $Y^n = X_1^n + X_2^n$ corrupted by erasures, and for the additive noise channel model $Y_n = X_1^n + X_2^n + Z^n$ for iid additive binary noise Z^n. Note that all alphabets and addition are over the binary field. Then, it can be shown that the multiple access capacities for both channels equal the rate region achieved by time-sharing between the points $(C, 0)$ and $(0, C)$, where $C = 1 - q(1)$ for the erasure model and $C = 1 - H(Z)$ for the additive noise model.

Note that time-sharing between two linear codes can also be described as a linear code, hence all points in the set of achievable rates are achievable by linear multiple access channel codes. For a sequence of rate-R single-transmitter, single-receiver channel codes, $\{b_n\}_{n=1}^\infty$, the multiple access

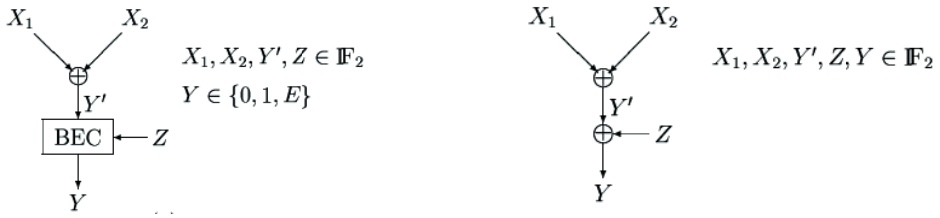

Figure 1: Binary additive multiple access channels with (a) erasures and (b) additive noise.

channel code achieving the $(\lambda, 1-\lambda)$ time-sharing solution between $(R, 0)$ and $(0, R)$ is a linear code with

$$[b^{\lambda}_{1,n}, b^{\lambda}_{2,n}] = \left(\begin{bmatrix} b_{\lambda n} & \mathbf{0}_{\lambda n \times (1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n \times \lambda nR} & \mathbf{0}_{(1-\lambda)n \times (1-\lambda)nR} \end{bmatrix}, \begin{bmatrix} \mathbf{0}_{\lambda n \times \lambda nR} & \mathbf{0}_{\lambda n \times (1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n \times \lambda nR} & b_{(1-\lambda)n} \end{bmatrix} \right).$$

Then, we decode the first λn channel outputs with the decoder for $\beta_{\lambda n}$ and the remaining outputs with the decoder for $\beta_{(1-\lambda)n}$.

With these mechanics, we obtain the channel coding theorems for both channel models. In particular, all rates (R_1, R_2) are achievable such that $R_1 + R_2 < 1 - q(1)$ for the erasure model, and $R_1 + R_2 < 1 - H(Z)$ for the additive noise model. Moreover, given these source and channel coding theorems, we have the following theorem of source-channel separation for our multiple access channels:

Theorem 7 *For the multiple access channel with erasures shown above, if source pair $(U_{1,1}, U_{2,1})$, $(U_{1,1}, U_{2,1})$, ... is drawn iid according to distribution $p(u_1, u_2)$ with $H(U_1, U_2) < 1 - q(1)$, then there exists a sequence of joint source-channel codes with probability of error tending to zero. Conversely, if $H(U_1, U_2) > 1 - q(1)$, then the probability of error for any communication system is bounded away from zero.*

A similar theorem is easy to formulate for multiple access channels with additive noise. Note that for a joint source-channel code, we can set up $n \times n$ matrices $C_{1,n}, C_{2,n}$ similarly as in the previous section, and for both channel models it can be shown that this scheme yields asymptotically negligible error probability for the sources with entropy lower than the channel capacities.

5 Broadcast Systems

A broadcast system source code consists of a single encoder and a collection of decoders. We consider a simple model with three receivers shown below. In this case, the linear encoder is a matrix of dimension

$$(\lceil nR_1 \rceil + \lceil nR_2 \rceil + \lceil nR_3 \rceil + \lceil nR_{12} \rceil + \lceil nR_{23} \rceil + \lceil nR_{13} \rceil + \lceil nR_{123} \rceil) \times n,$$

and we use typical set decoding. Then we have the following source coding theorem:

Theorem 8 *Consider samples of source vector $(U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123})$ drawn iid according to distribution $p = (u_1, u_2, u_3, u_{12}, u_{23}, u_{13}, u_{123})$ on $(\mathbb{F}_2)^7$. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of rate-$(R_1, R_2, R_3, R_{12}, R_{23}, R_{13}, R_{123})$ linear broadcast system source codes with coefficients chosen iid*

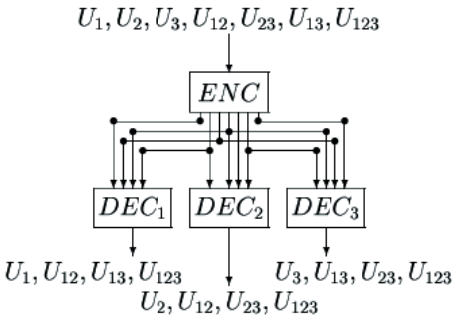

Figure 2: A broadcast system source code with three receivers.

Bernoulli(1/2). For any $s \subseteq \{1,2,3,12,23,13,123\}$, *let* $u_s = (u_a)_{a\in s}$, *and let* $(nR)_s = \sum_{a\in s}\lceil nR_a\rceil$. *Then for any rates satisfying*

$$\begin{aligned}(nR)_s &\geq H(U_s|U_{S_1-s}) \quad \forall s \subseteq S_1 = \{1,12,13,123\}, s \neq \emptyset\\ (nR)_s &\geq H(U_s|U_{S_2-s}) \quad \forall s \subseteq S_2 = \{2,12,23,123\}, s \neq \emptyset\\ (nR)_s &\geq H(U_s|U_{S_3-s}) \quad \forall s \subseteq S_3 = \{3,13,23,123\}, s \neq \emptyset,\end{aligned}$$

$\{A_n\}_{n=1}^{\infty}$ *achieves* $E[P_e(A_n)] \to 0$ *as* $n \to 0$.

Now we consider the following erasure broadcast channel models. In the model of Figure 3 (a),

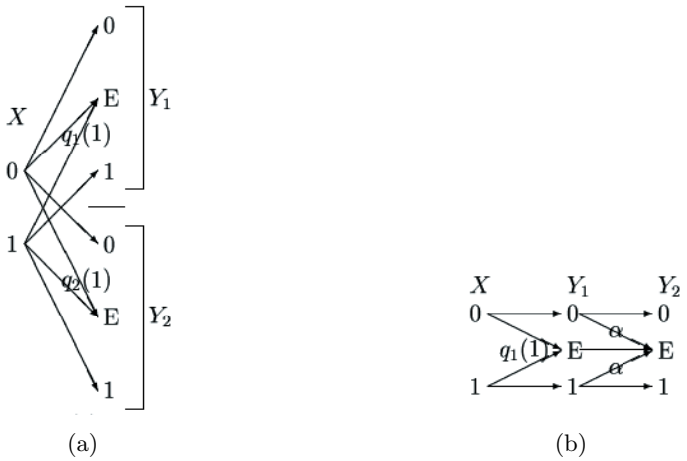

Figure 3: Binary additive multiple access channels with (a) erasures and (b) additive noise.

erasures are independent, but they are dependent in the model of Figure 3 (b). However, since the capacity of the broadcast channel depends only on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$, the capacity of the two channels are identical. The following Lemma proves that time-sharing to be optimal for broadcast coding over the given channels.

Lemma 1 *Consider a binary erasure channel with output alphabets {0,1,E} at each of two receivers. The erasure sequences* $Z_{1,1}$, $Z_{1,2}$, ... *and* $Z_{2,1}$, $Z_{2,2}$, ... *are drawn iid according to distributions* $q_1(z_1)$ *and* $q_2(z_2)$, *respectively. The joint distribution* $q(z_1,z_2)$ *can be arbitrary. The capacity region for sending independent information to the two receivers is*

$$\frac{R_1}{1-q_1(1)} + \frac{R_2}{1-q_2(1)} \leq 1.$$

For any achievable independent information rate pair (R_1, R_2)*, the rate triple* $(R'_1, R'_2, R'_{12}) = (R_1, R_2 - R_0, R_0)$ *with common information rate* R'_{12} *is also achievable for any* $R_0 < R_2$*.*

As mentioned before, time-sharing of linear channel codes yields also a linear code, linear broadcast channel codes on the erasure channel achieves the optimal rates.

However, there exist no results to prove the optimality of linear broadcast codes for the additive noise broadcast channel model. In this case, time-sharing is not optimal and direct application of the techniques used before fail to achieve the optimal performance.

6 Input-Dependent Noise

If the noise random variable is dependent on the channel input, we can show that the previous techniques do not apply. More precisely, separation does not hold in general for *linear* codes on single-transmitter, single-receiver channels with input dependent noise. However, there exists a more general example, i.e., not confined to linear codes, showing that source-channel separation fails owing to input-dependent noise in the case of additive multiple access channel with additive noise. This can be summarized as the following theorem:

Theorem 9 *Let* Z_1*,* Z_2*, ... be the noise random process, and* $X_{1,i}$*,* $X_{2,i}$ *be the channel inputs at time* i*. Then the channel output at time* i *is* $Y_i = X_{1,i} + X_{2,i} + Z_i$*. Separation fails when* Z_i *and* $(X_{1,i}, X_{2,i})$ *are statistically dependent random variables.*

7 The Case for End-to-End Coding

Traditional routing relies on simply repeat and forward techniques, but the network coding literature shows some simple multi-case examples where that approach fails in achieving the optimal performance. However, this network coding model may also fail due to the implicit assumption that source and channel coding are performed separately from network coding at the edges of the network.

Figure 4: Networks for which separation of (a) source-network or (b) channel-network coding fails

Consider a network in Figure 4 (a) where all links are lossless and the capacities are as shown in the figure. Let us assume that sources (U_1, U_2) are dependent random variables, with $H(U_1, U_2) < H(U_1) + H(U_2)$. Achieving reliable communication requires the descriptions received by nodes 1 and 2 to be dependent dependent random variables and requires sources U_1 and U_2 to be re-compressed at nodes 1 and 2, respectively. Hence, separation of source and network coding fails.

On the other hand, a network in Figure 4 (b) shows the case where separation of channel and network coding fails. The channel between node 0 and nodes 1 and 2 is a broadcast erasure channel with independent erasures of probabilities $q(1) = q(2) = q$, and the network between nodes 1 and

2 and node 3 is a multiple access channel without interference. If R_1 and R_2 are the capacities of the links to receivers 1 and 2, $R_1 + R_2 < 1 - q$ by Lemma 1. Now optimal network coding on the given multiple access channel gives a maximal rate of $1 - q$. However, if we do not force zero error probability between node 0 and nodes 1 and 2 but instead simply forward the information, then the capacity of resulting code is $1 - q^2$.

While the second example illustrates that failure to decode at intermediate nodes can yield superior performance, we can find another example showing the opposite, i.e., decoding at intermediate nodes is better. Hence, these examples suggest the need for end-to-end coding in network environments. Such failure of separation across canonical network systems entails a great challenge to optimal code design in networks.

References

[1] M. Effros, M. Médard, T. Ho, S. Ray, D. Karger, R. Koetter, and B. Hassibi, "Linear network codes: a unified framework for source, channel, and network coding," *Submitted to IEEE Transactions on Information Theory*, 2003.