Multilevel Broadcast Networks

Shashi Borade & Lizhong Zheng
LIDS & EECS
Massachusetts Institute of Technology
Cambridge, MA 02139
{spb,lizhong}@mit.edu

Mitchell Trott
Mobile & Media Systems Lab
Hewlett Packard Laboratories
Palo Alto, CA 92016
mitchell.trott@hp.com

***Abstract*— THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. We formulate a broadcast problem, where based on their quality of observations, outputs at various receivers are represented on a graph (called "degradation graph"). If receiver Z is a physically degraded version of receiver Y, then node Z is a child of node Y in this graph. This generalization of the classical degraded broadcast channel provides a framework for various situations where at least some information should be available to receivers with partial (or noisier) observations. Upper and lower bounds are obtained on the capacity region. The upper bound is based on auxiliary variables, whose structure is described by the mirror image of the channel's degradation graph. As a special case of our problem, a packet broadcast network is considered.**

I. Introduction

The conventional information theory for point-to-point communication focuses on the decoding of a block of data bits from the observation of a block of output symbols of the channel. An extension of this theory, which is useful in network communication systems, is to consider the data transmission when only different forms of *partial* observations are available at the receiver. It is natural to connect such problems to broadcast channels, where possible forms of partial observation are mapped to virtual receivers, each decoding some part of the data message. One difficulty of this approach is to describe the relation between these different partial observations, which are not necessarily degraded versions of one another. This gives rise to our formulation of a broadcast network.

A simple manner in which these partial observations could be related to each other is a Markov chain, which gives rise to the classical degraded broadcast channel.

$$X - Y_0 - Y_1 - \cdots - Y_{L-1} \tag{1}$$

For this channel with input X, amongst any two outputs Y_k and Y_m, one is a degraded version of the other. To get rid of this limitation, we model the interdependence between these partial observations using a "degradation tree" (see Figure 1 for example). This graph denotes that the network channel $P_{YW_1W_2Z_1Z_2|X}$ can be decomposed as follows

$$P_{YW_1W_2Z_1Z_2|X} = P_{Y|X}P_{W_1W_2|Y}P_{Z_1|W_1}P_{Z_2|W_2} \tag{2}$$

The root of this degradation tree denotes the input X, which is followed by a unique output Y. It is followed by its degraded versions W_1 and W_2, which in turn have their own degraded

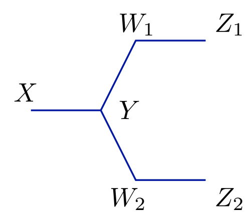

Fig. 1. An example of degradation graph.

versions Z_1 and Z_2. Note that here for example Z_2 need not be a degraded version of W_1.

The users in this degradation graph and their partial observations could have various physical interpretations. For example, it could represent users which are physically farther from the transmitter and hence receive noisier observations. It could also represent users which only listen during a fraction of the entire message transmission. As discussed before, even when there is only one physical user, the degradation graph could be used to represent the possible forms of partial observation.

One should be able to disseminate information such that even with partial observations, a user can receive some part of the transmitted information. Moreover, the information received should grow as the quality of observations improves. For example, in image transmission, even the weaker users would like to get a coarse image, whereas the stronger users would want some additional image resolution.

To formalize this notion of information dissemination, we can think of the degradation graph as a sequence of multiple layers, where layer k denotes the set of nodes at distance $k+1$ from the root. Outputs received at users in layer $k+1$ are degraded versions of outputs received in layer k. Assuming total L layers, let the total information be split into L parts. In our formulation, even the last layer users should get the first part of this information, the second last layer users should get the first as well as the second part and so on. Specifically, layer $L-1$ should be able to decode message M_{L-1}, layer $L-2$ should decode M_{L-2} as well as M_{L-1}. In general, layer k should decode $M_k, M_{k+1} \cdots M_{L-1}$. For example in Fig. 1, Z_1 and Z_2 are in layer 2, W_1 and W_2 are in layer 1 and Y is in layer 0. Hence both Z_1, Z_2 want message M_2; both W_1, W_2 want M_1 and M_2; and Y wants M_0, M_1 and M_2.

Previous works addressing similar concept of information dissemination include multilevel diversity coding [2] and priority encoded transmission [4].

We assume that messages $M_0, M_2 \cdots M_{L-1}$ are mutually independent. Message M_k $(0 \leq k \leq L-1)$ is chosen uniformly from $\{1, 2 \dots, 2^{nR_k}\}$, where R_k denotes the rate of message k and n denotes the block length. We want to characterize the achievable rate region $(R_0, R_2 \dots, R_{L-1})$.

We define degradation graphs to include slightly greater generality than the trees illustrated in Figure 1. A degradation graph could have a node in layer k which has two parents in layer $k-1$, provided it is a deterministic function of each parent.

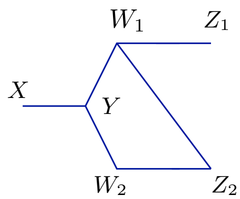

Fig. 2. A degradation graph which is not tree.

For example in Figure 2, Z_2 has two parents W_1 and W_2, which means $Z_2 = f_1(W_1) = f_2(W_2)$ where f_1 and f_2 are deterministic functions. This generalization of allowing multiple parents is useful for modeling packet broadcast networks for example. One can now think of the degradation graph as a directed acyclic graph, where an edge directed from node A to node B denotes that either B is a noisier version of A or B is a deterministic function of A.

The remaining paper is organized as follows. Section II studies the classical physically degraded broadcast channel. A converse was proved for the case of two users in [1] and also [5]. We first prove the converse for $L > 2$ users using a method similar to [1] for two users. This proof is also useful for the converse for a general degradation graph. As an example, we consider broadcast over binary erasure channels. A simple scheme of time-sharing between some binary linear codes can achieve the capacity region of this channel.

A channel with arbitrary degradation graph is studied in Section III. An achievable rate region is presented and a converse is proved which is based on mirror image of the degradation graph. This converse gives the capacity region for a class of degradation graphs. As a special case of our problem, a packet broadcast network is studied in Section 4, where our achievable rate region calculated in closed form. This example shows how various problems such as priority encoded transmission and multilevel diversity coding ([2] and [4]) can be analyzed using degradation graphs.

II. Classical degraded broadcast channel with multiple receivers

In a physically degraded broadcast channel to L users (Y_0, $Y_1 \dots Y_{L-1}$), the network channel can be decomposed as

$$P_{Y_0 Y_1 \cdots Y_{L-1}|X} = P_{Y_0|X} P_{Y_1|Y_0} \cdots P_{Y_{L-1}|Y_{L-2}} \tag{3}$$

Thus this channel is fully described by the L probability transition functions $P_{Y_0|X}$, $P_{Y_1|Y_0} \cdots P_{Y_{L-1}|Y_{L-2}}$ for this channel. The following achievable region was proved in [6], [7] using superposition codes.

Consider a Markov chain

$$U_{L-1} - U_{L-2} \cdots - U_1 - X - Y_0 - Y_1 \cdots - Y_{L-1} \tag{4}$$

where $U_{L-1}, \cdots U_1$ are called auxiliary random variables. The joint distribution of all variables above is given by

$$(Q_{U_{L-1}U_{L-2}\cdots U_1 X}) \cdot (P_{Y_0 Y_1 \cdots Y_{L-1}|X})$$

where the second term is described by the channel as in (3). The first term denotes the joint distribution of input X and auxiliary variables. It is chosen by the code designer and has a Markov structure as follows. Throughout this paper, we will use Q for distributions chosen by the code designer and P for channel transition probabilities.

$$Q_{U_{L-1}U_{L-2}\cdots U_1 X} = Q_{U_{L-1}} \left(\prod_{k=L-1}^{2} Q_{U_{k-1}|U_k} \right) Q_{X|U_1} \tag{5}$$

Then rate tuples $(R_0, R_1 \cdots R_{L-1})$ satisfying following inequalities are achievable,

$$\begin{aligned} R_{L-1} &\leq I(U_{L-1}; Y_{L-1}) \\ R_k &\leq I(U_k; Y_k | U_{k+1}) \text{ for } k \in [1 : L-2] \\ R_0 &\leq I(X; Y_0 | U_1) \end{aligned}$$

Let $\mathcal{R}_Q$ denote this rate region for a particular choice of Q satisfying (4) and (5). The achievable region is obtained *conv* $\left(\bigcup_{Q \in \text{Markov}} \mathcal{R}_Q \right)$ where the union is taken over all joint distributions Q with a Markov structure in (4,5) and *conv*$(\cdot)$ denotes the convex hull operation (arising from time-sharing arguments).

One may wonder why is it necessary that the auxiliary variables should have a Markov structure. The code designer could have chosen any joint distribution $Q_{U_{L-1}U_{L-2}\cdots U_1 X}$, which need not have a Markov structure. Using similar random coding construction as [7], one can show that rate-tuples obeying following equation are achievable with superposition.

$$\begin{aligned} R_{L-1} &\leq I(U_{L-1}; Y_{L-1}) \\ R_k &\leq I(U_k; Y_k | U_{k+1} U_{k+2} \cdots U_{L-1}) \ , \ k \in [1 : L-2] \\ R_0 &\leq I(X; Y_0 | U_1 U_2 \cdots U_{L-1}) \end{aligned}$$

We need to slightly modify the coding scheme in [7] for achieving the rates above. Now the distribution of the U_k codebook for user Y_k depends on all previous auxiliary codewords (in $U_{k+1}, U_{k+2} \cdots U_{L-1}$) for users $Y_{k+1}, Y_{k+2} \cdots Y_{L-1}$. In [7], this only depended on auxiliary U_{k+1} codeword for user Y_{k+1}. Similarly, now the distribution of X codeword for user Y_0 depends on all auxiliary codewords instead of just the auxiliary U_1 codeword for Y_1.

Let the achievable region above be denoted by $\mathcal{R}_Q^*$, where Q could be any joint distribution of auxiliaries and input. Since mutual information can decrease or increase by conditioning, it is unclear whether the convex hull *conv* $\left(\bigcup_Q \mathcal{R}_Q^* \right)$ over all joint distributions Q remains unchanged when Q is restricted to have a Markov structure. Next subsection clarifies this and proves the optimality of Markov structure.

A. *Converse for Degraded Broadcast Channels*

Now let us prove optimality of this achievable region on similar lines of Gallager's proof for the two-user case[1] [1]. This proof is helpful when proving the converse for general degradation graphs.

First we define the following function over non-zero vectors $\lambda \equiv (\lambda_1, \lambda_2, \cdots, \lambda_{L-1}) \geq 0$,

$$C(\lambda) = \sup_{Q \in \text{Markov}} \left(I(X;Y_0|U_1) + \sum_{k=1}^{L-1} \lambda_k I(U_k;Y_k|U_{k+1}) \right) \quad (6)$$

where the last term in the summation, $I(U_{L-1};Y_{L-1}|U_L)$, is a shorthand for $I(U_{L-1};Y_{L-1})$. The supremum is over all Markov chains $U_{L-1} - U_{L-2} \cdots U_1 - X$. Similar to [1], it can be shown that the supremum above is unchanged even if we restrict the cardinality of each U_i to that of X.

We will show that an achievable rate-tuple must satisfy

$$R(\lambda) \equiv R_0 + \sum_{k=1}^{L-1} \lambda_k R_k \leq C(\lambda) \qquad \forall\ \lambda \geq 0 \quad (7)$$

More specifically, we show that if a rate-tuple disobeys the bound above for any $\lambda \geq 0$, then vanishing error probability cannot be achieved for all users. By convex programming, this is the same as showing that any point outside the achievable region *conv* $\left(\bigcup_{Q \in \text{Markov}} \mathcal{R}_Q\right)$ is not achievable, which proves the optimality of the achievable region in [7]. The proof mainly follows from the following lemma.

Lemma 1: Let $Y_k^{1:n}$ denote a shorthand for all outputs at user Y_k from time 1 to n.

$$\begin{aligned} nC(\lambda) \;&\geq\; I(M_0;Y_0^{1:n}|M_1M_2\cdots M_{L-1}) && (8)\\ &+\; \sum_{k=1}^{L-2} \lambda_k I(M_k;Y_k^{1:n}|M_{k+1}M_{k+2}\cdots M_{L-1}) && (9)\\ &+\; I(M_{L-1};Y_{L-1}^{1:n}) && (10) \end{aligned}$$

Choice of auxiliaries: The proof of the lemma is omitted but its main component is our substitution of auxiliary random variable U_k at time j. It is defined as the set of

1) Past symbols (up to time $j-1$) observed at user Y_k as well as the past symbols observed at its physically degraded users (i.e., $Y_{k+1}, Y_{k+2}, \cdots Y_{L-1}$) and
2) the messages for layer i and all further layers, i.e., $M_k M_{k+1} \cdots M_{L-1}$.

This method for choice of auxiliaries is also used in proving our converse for general degradation graphs.

Theorem 2: For some $\lambda \geq 0$, $\epsilon > 0$, if a rate-tuple satisfies

$$R(\lambda) \quad \geq \quad \epsilon + C(\lambda) \quad (11)$$

then error probability cannot vanish for all users, because error probabilities $(p_0, p_1 \cdots p_{L-1})$ of the L receivers satisfy

$$(1 + p_0 n R_0) + \sum_{k=1}^{L-1} \lambda_k (1 + p_k n R_k) \geq n\epsilon \quad (12)$$

[1]It should be mentioned that we could not extend the converse proof in [5] for more than two users.

Proof of Theorem 2: If (11) holds, then by Lemma 1,

$$\begin{aligned} nR(\lambda) \;&\geq\; n\epsilon + I(M_0;Y_0^{1:n}|M_1M_2\cdots M_{L-1}) && (13)\\ &+\; \sum_{k=1}^{L-2} \lambda_k I(M_k;Y_k^{1:n}|M_{k+1}M_{k+2}\cdots M_{L-1}) && (14)\\ &+\; I(M_{L-1};Y_{L-1}^{1:n}) && (15) \end{aligned}$$

But each mutual information above can be bounded as:

$$\begin{aligned} & I(M_k;Y_k^{1:n}|M_{k+1}M_{k+2}\cdots M_{L-1})\\ \geq\;& H(M_k|M_{k+1}M_{k+2}\cdots M_{L-1}) - H(M_k|Y_k^{1:n})\\ =\;& nR_k - H(M_k|Y_k^{1:n})\\ \geq\;& nR_k - (1 + p_k n R_k) \end{aligned}$$

where the first inequality follows since conditioning reduces entropy, the second equality is due to independence of messages and the last step is due to Fano's inequality. Rearranging this and substituting back (13) yields (12). QED.

B. *Binary Erasure Channels*

Consider this simple achievable scheme for L binary erasure channels with erasure probabilities $e_1, e_2 \cdots e_L$. The block length n is divided into L separate blocks, where block k has length $n\alpha_k$. Receiver k's message of nR_k information bits is converted to $n\alpha_k$ coded bits. This can be done with a linear operation $r = Ab$, where b denotes a vector of the nR_k information bits, r is the vector of $n\alpha_k$ coded bits and A is a $n\alpha_k \times nR_k$ generator matrix.

We can choose each entry of A independently with uniform binary distribution. As block length grows large, essentially any nR_k rows of this matrix will be linearly independent with high probability. Thus with high probability, receiving any nR_k elements of r is sufficient to decode b. Now note that essentially $(1 - e_k) n\alpha_k$ coded bits will reach unerased at receiver k. Hence if $\alpha_k(1 - e_k) > R_k$, then with high probability, receiver k can decode its message of nR_k bits. Any receiver with smaller erasure probability will also decode this message. Since sum of α_k's is at most unity,

$$\sum_{k=1}^{L} R_k/(1 - e_k) < \sum_{k=1}^{L} \alpha_k \leq 1$$

This indeed is the capacity region for this broadcast channel. Thus a simple scheme of dividing the block-length amongst random linear codes achieves the capacity region here.

III. Achievability and Converse for General Degradation Graphs

Consider an arbitrary degradation graph (such as Figure 1 or Figure 2). Recall that layer k denotes all nodes at distance $k+1$ from the root node of input X. All nodes in layer k of this degradation graph are interested in $M_k, M_{k+1} \cdots M_{L-1}$, where L is the depth of the graph from the root node. The rate of message M_i is denoted by R_i.

We can use similar random coding construction as [7], which is based on superposition. Using standard typicality arguments, we can prove the following achievable rate region.

For concreteness and clarity, we will state our the result for the particular degradation graph in Figure 1. Similar rate-region can be written down for any degradation graph.

As the degradation graph in Figure 1 has $L = 3$ layers, we choose $L - 1 = 2$ auxiliary random variables U_1, U_2 such that joint distribution of all auxiliaries, input X and all outputs has the following Markov structure.

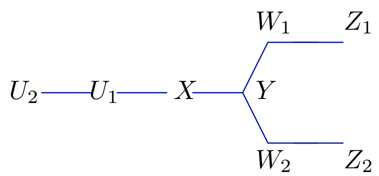

Fig. 3. Markov structure for achievability

Thus the joint distribution of all variables is given by

$$Q_{U_2 U_1 X} P_{Y W_1 W_2 Z_1 Z_2 | X}$$

where the second term is determined by the network channel as in (2). The first term is chosen by the code designer and satisfies the Markov structure in above figure.

Theorem 3: For every choice of $Q_{U_2 U_1 X}$ consistent with the Markov structure above, rate-tuples obeying following conditions are achievable:

$$\begin{aligned} R_2 &\leq \min\left(I(U_2; Z_1), I(U_2; Z_2)\right) \\ R_1 &\leq \min\left(I(U_1; W_1 | U_2), I(U_1; W_2 | U_2)\right) \\ R_0 &\leq I(X; Y | U_1) \end{aligned}$$

Let $\mathcal{R}_Q$ denote this region where Q is joint distribution of (U_1, U_2, X). Then any rate-tuple in *conv*$(\bigcup_{Q \in \text{Markov}} \mathcal{R}_Q)$ is achievable where Q denotes all distribution satisfying the Markov structure in Figure 3.

Similar to Section II, one may wonder why restrict to a Markov chain for auxiliaries. We can indeed choose $Q_{U_2 U_1 X}$ which is not a Markov chain $U_2 - U_1 - X$. By similar arguments as in Section II, we get the following achievable rate region, called as $\mathcal{R}^*_Q$, for every distribution Q on (U_2, U_1, X).

$$\begin{aligned} R_2 &\leq \min\left(I(U_2; Z_1), I(U_2; Z_2)\right) \\ R_1 &\leq \min\left(I(U_1; W_1 | U_2), I(U_1; W_2 | U_2)\right) \\ R_0 &\leq I(X; Y | U_1 U_2) \end{aligned}$$

That is, mutual information between U_k and an output in layer k is conditioned on $U_{k+1} U_{k+2} \cdots U_{L-1}$ instead of just on U_{k+1} as in the case of a Markov chain.

Theorem 4: An achievable region is given by *conv*$(\bigcup_Q \mathcal{R}^*_Q)$, where the union is over all joint distributions of auxiliaries and input X.

However, we believe that similar to the classical degraded broadcast channel in Section II, Markov chain of auxiliaries is optimal for superposition coding and achieves the entire capacity region.

Conjecture 5: We conjecture that superposition coding with a Markov chain of auxiliary variables is optimal. That is, achievable region in Theorem 3, i.e. *conv*$(\bigcup_{Q \in \text{Markov}} \mathcal{R}_Q)$, equals the capacity region.

Next section provides a converse for the rate-region, which verifies the conjecture for a class of degradation graphs. However, it does not prove the conjecture for a general degradation graph. Note that our problem formulation ensures that a node (say A) in layer k can decode all the messages for any node in further layers (say B). This holds true even if B is not a degraded version of A. For example, Z_2 is not a degraded version W_1 in Fig. 1, but W_1 can decode what Z_1 can decode. We believe this property can be used to prove our conjecture using ideas from [8].

A. Mirror Image Converse

Using the same guideline for the *choice of auxiliaries* as discussed in Section II, we can obtain the following upper bound on the achievable rate region.

We consider the degradation graph in Figure 2 for concreteness and converse for a general degradation graph can be expressed similarly. Furthermore, for better clarity, lets convert the degradation graph in Figure 2 to a directed graph, where a directed edge from node A to node B denotes that B is degraded version of A. For example, an edge directed from Y to W_1 represents that W_1 is a degraded version of Y. If multiple edges are directed towards a node (e.g. Z_2), then it is a deterministic function of all its parents (e.g. W_1 and W_2).

With this setup up we are ready to state the converse. Given the degradation graph, create its mirror image and attach it behind the root node X as follows:

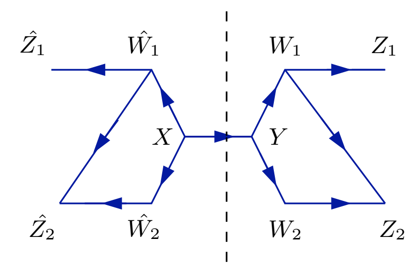

Fig. 4. Upper bound with mirror image structure of auxiliaries

The auxiliary variables $(\hat{Z}_1, \hat{Z}_2, \hat{W}_1, \hat{W}_2)$ thus have the same degradation graph as the channel outputs (Z_1, Z_2, W_1, W_2). For example, $\hat{Z}_1$ is a degraded version of $\hat{W}_1$, whereas $\hat{Z}_2$ is a deterministic function of both $\hat{W}_1$ and $\hat{W}_2$. For a distribution Q on auxiliaries satisfying the above mirror structure, let $\bar{\mathcal{R}}_Q$ denote the following rate-region.

$$\begin{aligned} R_2 &\leq \min\left(I(\hat{Z}_1; Z_1), I(\hat{Z}_2; Z_2)\right) \\ R_1 &\leq \min\left(I(\hat{W}_1; W_1 | \hat{Z}_1 \hat{Z}_2), I(\hat{W}_2; W_2 | \hat{Z}_2)\right) \\ R_0 &\leq I(X; Y | \hat{W}_1 \hat{W}_2) \end{aligned}$$

Here each mutual information between a channel output and its mirror image auxiliary variable is conditioned on all the children of the mirror auxiliary variable in the mirror structure.

Theorem 6: Every achievable rate-tuple is in the convex hull *conv*$(\bigcup_{Q \in \text{Markov}} \bar{\mathcal{R}}_Q)$, where the union is over all distributions of auxiliaries which satisfy the mirror structure.

Note that converse for the classical degraded broadcast channel in Section II is a special case of this converse because mirror image of a straight line is a straight line. This incidentally proved optimality of the achievable rates in [7], where a Markov chain of auxiliaries is used for superposition coding.

Moreover, using this converse and a simple property of mutual information, we can prove Conjecture 5 for a class of degradation graphs. This class of graphs is where all outputs in layer k are children of every node in previous layer (see Figure 5). In this case, edges between two adjacent layers form a complete bi-partite graph. Nonetheless, for an arbitrary degradation graph, our converse may not match our achievable region because mutual information could either decrease or increase by conditioning.

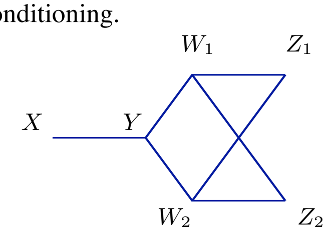

Fig. 5. A degradation graph for which Conjecture 5 is true.

IV. Packet Broadcast Networks

As an application of the degradation graph framework, we consider a situation where transmitter emits a fixed number L of n-bit packets. Each of these packets can be either perfectly received or be erased completely. A user can receive any of the $2^L - 1$ possible subsets of these L packets. We want to ensure that any user who gets k number of packets can decode the k messages $(M_{L-k}, M_{L-k+1}, \cdots M_{L-1})$. Earlier problems of multilevel diversity coding [2], [3] and priority encoded transmission [4] can be modeled this way. Degradation graphs provide a common framework to address them. In addition, this framework can be also used to model errors in packets.

Let us assume $L = 3$ packets are transmitted for simplicity, although similar analysis can be performed for any L. The degradation graph in this situation is shown below. The actual transmission is denoted as (X_1, X_2, X_3) where each X_i represents a bit from packet i. Various receivers receive all seven subsets of this transmission. For better clarity, X_{123} is used to denote $(X_1X_2X_3)$ for example.

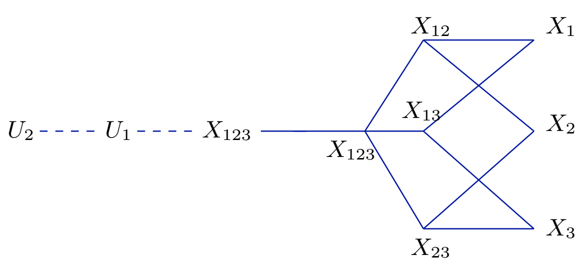

Fig. 6. Solid line shows the degradation graph for packet broadcast network. Dotted line and U_1, U_2 are the auxiliaries for achievability in Theorem 3.

Degradation graphs can be also used to model asymmetric situations, where only certain subsets of transmitted packets can be received. For example, there may not be any user who gets packets 2 and 3. This can be modeled by simply removing X_{23} from the degradation graph. These asymmetric situations seem particularly relevant for distributed storage [3]. In a distributed storage system with L storage locations, each user may have access to a subset of these locations. Moreover, now the effect of errors in stored data can also be analyzed using degradation graphs.

Now let us calculate the achievable region for Fig. 6 from Theorem 3. We show that this achievable rate-region is given by $R_2 + R_1/2 + R_0/3 \le 1$, where R_k is the rate of message M_k for layer k. From Theorem 3,

$$R_2 \le I(U_2;X_1),\ R_2 \le I(U_2;X_2) \quad \& \ R_2 \le I(U_2;X_3) \quad (16)$$

$$R_1 \le I(U_1;X_{12}|U_2) \quad (17)$$

$$R_1 \le I(U_1;X_{13}|U_2) \quad (18)$$

$$R_1 \le I(U_1;X_{23}|U_2) \quad (19)$$

$$R_0 \le I(X_{123};X_{123}|U_1) = H(X_{123}|U_1) \quad (20)$$

Adding inequalities in $(16)\times 2 + (17) + (18) + (19) + (20)\times 2$ implies $6R_2 + 3R_1 + 2R_0$ is not greater than

$$\left[2\sum_{i=1}^{3} H(X_i)\right] - \left[2\sum_{i=1}^{3} H(X_i|U_2) - \sum_{i\neq j} H(X_{ij}|U_2)\right] + \left[2H(X_{123}|U_1) - \sum_{i\neq j} H(X_{ij}|U_1)\right]$$

First bracket above is at most 6, second bracket equals $\sum_{i\neq j} I(X_i;X_j|U_2)$ and hence non-negative. Third bracket equals $-I(X_3;X_1|X_2,U_1) - I(X_2;X_{13}|U_1)$, which is non-positive. Thus $6R_2 + 3R_1 + 2R_0 \le 6$, proving our achievable region. This inequality can be achieved with equality if X_i's are independent conditioned on U_1 and pairwise independent conditioned on U_2.

For any K number of packets, similar analysis can be done to prove that achievable region in Theorem 3 equals $\sum_{i=0}^{K-1} R_i/(K-i) \le 1$. In fact, this was shown to be the entire capacity region in [2], [4]. Thus our conjecture is verified in this problem.

Acknowledgements

We thank Robert Gallager for his many helpful comments.

References

[1] R. Gallager, Capacity and coding for degraded broadcast channels, *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3-14, 1974.
[2] R. Yeung and Z. Zhang, “On symmetrical multilevel diversity coding,” *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 609-621, Mar. 1999.
[3] J. R. Roche, Distributed information storage, *Ph.D. dissertation*, Stanford University, Mar. 1992.
[4] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, Priority encoding transmission, *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737-1744, Nov. 1996.
[5] R. Ahlswede and J. Korner, “Source coding with side information and a converse for degraded broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 21, no. 6, pp. 629-637, Nov. 1975.
[6] T. M. Cover, ”Broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 18, pp. 2-14, Jan 1972.
[7] P. Bergmans, ”Random coding theorem for broadcast channels with degraded components,” *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197-207, Mar 1973.
[8] K. Marton and J. Korner, “ General broadcast channels with degraded message sets ,” *IEEE Trans. Inform. Theory*, vol. 23, no. 1, pp. 60-64, Jan. 1977.