6.962: Week 3 Summary of Discussion

Presenter: Brett Schein

Date: September 27, 2000

Paper: "Capacity Theorems for the Relay Channel",
T. Cover and A. El Gamal, IEEE IT, Sept. 1979.

I. Introduction

It is well known that, in some contexts, cooperation between terminals in a multiple terminal system can enlarge the set of reliably achievable communication rates. This is true, for example, when terminals obtain information about what other terminals are sending through feedback [5], [8] or through side channels [4].

For systems where power is of primary importance, such as in wireless or ad hoc networks, terminals can cooperate by sending signals with a common component. This common component coherently combines at a receiver, resulting in an increased effective power. Exploiting this requires common information in some form at distributed points within a network. As an example, one such technique has recently been suggested for reducing the deleterious effects of multipath fading in a particular wireless multiple access system [9]. In a wireless environment, such coherent combination would require synchronism between the *carriers* rather than symbols, thus minimally requiring some form of feedback or common timing information. While this level of precision may be difficult to achieve in a real, distributed system, it is important to investigate its possibility for improving both real-world systems and theoretical understanding of networks.

To this end, we develop simple mathematical models that allow an exploration (and hopefully an elucidation) of the fundamental issues while eliminating all but the most crucial complexity. We can build later upon our understanding. One such simple, applicable model called the relay channel was defined by Van der Meulen [3]. This network is pictured in Figure 1. The relay channel consists of an input terminal with input $\mathbf{X_1}$, a relay terminal which observes $\mathbf{Y_1}$ and transmits $\mathbf{X_2}$, and an ultimate receiver terminal which observes $\mathbf{Y}$. The only source of extrinsic information is the input terminal, and the only purpose of the relay terminal is to help get information to the receiver. The relay is an arbitrarily complex processor whose transmission at time k, $\mathbf{X}_{2,k}$, can depend upon all observations received up to but not including time k, $\mathbf{Y_1}^{k-1}$. In the literature, the relay channel is most often a

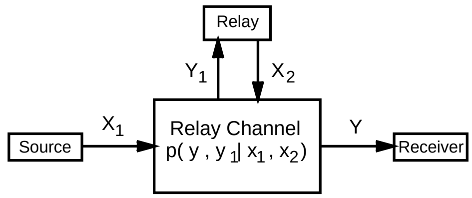

Fig. 1. The relay channel

discrete time, finite alphabet channel which is defined by the abstract alphabets and a probability transition matrix $p(y_1, y|x_1, x_2)$. The goal is to find the information-theoretic capacity of the network. The solution is still unknown in all but degenerate cases.

We will focus in detail on the work of Cover and El Gamal, "Capacity Theorems for the Relay Channel." This work contains a number of interesting approaches to the problem, and we will proceed serially through most of their results. The capacity is found in several degenerate cases. For directly complimentary literature, see [6], [1], [7], [10]. All of these address special cases where one or both of the observations are deterministic functions of the inputs. For the particularly brave of heart, see [11].

We feel that a better model for basic understanding of noisy networks is that pictured in Figure 2. Here the direct path from source to receiver in the relay channel is replaced with a second relay terminal. The core issue to be resolved in such systems is how to design the cooperation amongst multiple terminals in a distributed system. Here, the second relay terminal allows a (discrete time) synchronization in transmitting to the receiver — eliminating the need for the analytically interesting but complicating approach of superposition coding exploited by Cover and El Gamal. These more complicated issues need to be addressed for general mesh networks (acyclic or otherwise), but this will more easily be explored once we understand the basic issues. Hopefully we will have time to discuss this latter network a little at the end of the discussion.

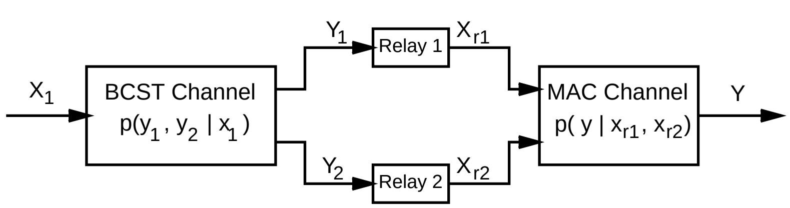

Fig. 2. The double relay network

II. The Paper

Cover and El Gamal begin by defining a degenerate class of relay channels. The relay channel is physically degraded if $p(y_1, y|x_1, x_2) = p(y_1|x_1, x_2) \cdot p(y|y_1, x_2)$ for all possible $\{x_1, x_2, y_1, y\}$. This means that the receiver terminal's observation $\mathbf{Y}$ is in some sense a noisy version of the *experimental outcome* y_1 (specifically, $\mathbf{X_1} \rightarrow (\mathbf{X_2}, \mathbf{Y_1}) \rightarrow \mathbf{Y}$ form a Markov chain). This seems like an extremely uninteresting class except as it applies to perfect feedback channels. The more interesting class is the stochastically degraded relay channel which satisfies $p(y|x_1, x_2) = \sum_{y_1} p(y_1|x_1, x_2) \cdot t(y|y_1, x_2)$ for some transition matrix $t(\cdot)$. We may also rather naturally assume that the observations $\mathbf{Y_1}$ and $\mathbf{Y}$ are conditionally independent when conditioned on the inputs $\mathbf{X_1}$ and $\mathbf{X_2}$, so that $p(y_1, y|x_1, x_2) = p(y_1|x_1, x_2) \cdot p(y|x_1, x_2)$. For the stochastically degraded relay channel, the receiver observation $\mathbf{Y}$ looks like a noisier but statistically independent observation of the input signals $(\mathbf{X_1}, \mathbf{X_2})$.

As a side note, a physically degraded channel is easier to deal with analytically — in particular in proving converse theorems. For example, a simple converse proof for the stochastically degraded broadcast channel proceeds in two steps. First, we show that the capacity (region) depends only on the marginal distributions from the source to the observations. Then we prove a converse for the physically degraded counterpart (see [2, pp. 454-5, Ex. 10, 11]). For the degraded relay channel, the same procedure does not work.

That being noted, the capacity of the physically degraded relay channel is given by

$$\mathbf{C} = \max_{p(x_1, x_2)} [\, \min\, [\, I(\mathbf{X_1}, \mathbf{X_2}\, ; \mathbf{Y}),\ I(\mathbf{X_1}; \mathbf{Y_1}|\, \mathbf{X_2})\,]\,]\ . \tag{1}$$

The achievability argument holds for the general relay channel. It involves a two-stage block Markov coding scheme. In the first stage, the relay learns the input message exactly, while the receiver forms a list of possible messages consistent with its observation y. In the second stage, the relay and the source cooperate to resolve the remaining uncertainty at the receiver. The two stages are superimposed so that stage 2 of iteration k and stage 1 of iteration $k+1$ occur simultaneously.

The converse for the physically degraded channel follows by first proving a converse for the general relay channel,

$$\mathbf{C} \leq \max_{p(x_1, x_2)} [\, \min\, [\, I(\mathbf{X_1}, \mathbf{X_2}\, ; \mathbf{Y}),\ I(\mathbf{X_1}; \mathbf{Y_1}, \mathbf{Y}|\, \mathbf{X_2})\,]\,]\ . \tag{2}$$

The two terms in (2) are essentially applications of the data processing theorem. They can be viewed as information-theoretic versions of cut-sets through the network, reminiscent of the discussion of the first week. The converse for the physically degraded channel now follows by observing that

$H(\mathbf{X_1}|\ \mathbf{X_2}, \mathbf{Y_1}, \mathbf{Y}) = H(\mathbf{X_1}|\ \mathbf{X_2}, \mathbf{Y_1})$ so that $I(\mathbf{X_1}; \mathbf{Y_1}, \mathbf{Y}|\ \mathbf{X_2}) = I(\mathbf{X_1}; \mathbf{Y_1}|\ \mathbf{X_2})$. This reflects the fact that the physically degraded observation $\mathbf{Y}$ yields no useful information about the source $\mathbf{X_1}$ once we have access to the better observation $\mathbf{Y_1}$. This is fundamentally why physically degraded channels are easier to handle analytically, as well as why this converse would not necessarily be tight for the stochastically degraded channel. Indeed, for the double relay network of Figure 2, most reasonable network models (e.g., the Gaussian, power-limited version) can achieve rates larger than the appropriate version of (1) when the multi-access (MAC) side is not too weak.

Next, Cover and El Gamal essentially evaluate (1) for the physically degraded Gaussian channel with average power constraints, though a little care is needed in doing so. It can be seen that to maximize the first term in (1), we want to make the input symbols $\mathbf{X_1}$ and $\mathbf{X_2}$ highly correlated, i.e., to cooperate closely between the input and relay terminals. On the other hand, to maximize the second term in (1), we want to make the input symbols $\mathbf{X_1}$ and $\mathbf{X_2}$ independent. These are conflicting objectives. The authors find a way to track the relevant relationship between $\mathbf{X_1}$ and $\mathbf{X_2}$, tying together the second moments via $(\mathbb{E}\{\mathbf{X_1}|\ \mathbf{X_2}\})^2$. An essentially identical approach was later used by Ozarow in deriving the converse to the Gaussian multi-access channel with feedback [8].

The authors then proceed with two theorems (Theorems 6 and 7) which exploit an alternative approach to a two-stage coding process. We will describe only Theorem 6 since Theorem 7 is the simultaneous combination of the approaches of 1 and 6 (and results in a more general achievability statement which is quite difficult to assess). Rather than decoding the message, the relay sends an estimate $\widehat{\mathbf{Y}}_\mathbf{1}$ of its observation $\mathbf{Y_1}$. As before, the receiver uses a two-stage decoding process. In the first stage, the receiver forms a list of possible estimates $\widehat{\mathbf{Y}}_\mathbf{1}$ consistent with its observation y. In the second stage, the receiver resolves the estimate $\hat{y}_1$ and then decodes to the unique message consistent with the pair of observations $(\hat{y}_1, y)$.

The simplest way to view this procedure is to view the estimate possibilities $\{\widehat{\mathbf{Y}}_\mathbf{1}\}$ as a rate-distortion codebook for the "observation source" $\mathbf{Y_1}$. For efficiency, the rate-distortion codewords are randomly combined into bins. The appropriate bin index is sent as a noisy channel codeword for the single-user channel $\mathbf{X_2} \rightarrow \mathbf{Y}$ (with $\mathbf{X_1}$ being treated as independent noise). The end-to-end achievability result follows like a point-to-point, single-user channel coding argument with input $\mathbf{X_1}$ and a pair of observations $(\widehat{\mathbf{Y}}_\mathbf{1}, \mathbf{Y})$. The achievability theorem, Theorem 6, states that

$$C \geq \max I(\mathbf{X_1}; \widehat{\mathbf{Y}}_\mathbf{1}, \mathbf{Y} |\ \mathbf{X_2}) \tag{3}$$

for any single-letter distribution that satisfies

$$I(\mathbf{X_2}; \mathbf{Y}) \geq I(\mathbf{Y_1}; \widehat{\mathbf{Y}}_\mathbf{1} |\ \mathbf{X_2}, \mathbf{Y}), \tag{4}$$

where $p(x_1, x_2, y_1, y_2, \hat{y}_1) = p(x_1)p(x_2)p(y, y_1|x_1, x_2)p(\hat{y}_1|y_1, x_2)$.

The inequality (4) reflects the requirement for the rate-distortion and binning approach to work. No explicit cooperation between the relay and the input terminal is used in the code construction, and this is reflected in the indepedent distribution $p(x_1, x_2) = p(x_1) \cdot p(x_2)$.

III. Final remarks

As a final note, consider the symmetric Gaussian double relay network of Figure 3. Here we assume that the noise powers on the left-hand side are the same, i.e., $\mathrm{N}_{Z_1} = \mathrm{N}_{Z_2}$. Further, we assume the input has an average power constraint P_{X_1}, and the relays have the same average power constraint $\mathrm{P}_{X_{r1}} = \mathrm{P}_{X_{r2}}$ (different from P_{X_1}). We normalize the symmetric network parameters by the signal-to-noise ratios $\mathrm{S_{in}} = \frac{\mathrm{P}_{X_1}}{\mathrm{N}_{Z_1}} = \frac{\mathrm{P}_{X_1}}{\mathrm{N}_{Z_2}}$ and $\mathrm{S_{rel}} = \frac{\mathrm{P}_{X_{r1}}}{\mathrm{N}_Z} = \frac{\mathrm{P}_{X_{r2}}}{\mathrm{N}_Z}$.

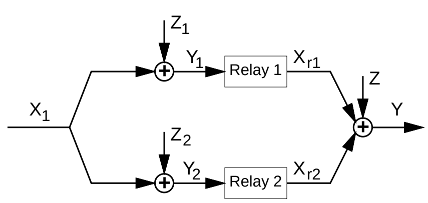

Fig. 3. The Gaussian double relay network

Consider using the relays as simple amplifiers (transponders), so that

$$\mathbf{X}_{r1} = \alpha \cdot \mathbf{Y_1},$$

$$\mathbf{X}_{r2} = \alpha \cdot \mathbf{Y_2},$$

$$\alpha = \sqrt{\frac{\mathrm{P}_{X_{r1}}}{\mathrm{P}_{X_1} + \mathrm{N}_{Z_1}}}\ .$$

It can be shown by single-user, point-to-point techniques that we can achieve communication rates

$$\begin{aligned} R &= I(\mathbf{X_1}; \mathbf{X}_{r1} + \mathbf{X}_{r2} + \mathbf{Z}) \\ &= \frac{1}{2}\log_2\left(1 + \frac{4\mathrm{S_{rel}}\mathrm{S_{in}}}{1 + 2\mathrm{S_{rel}} + \mathrm{S_{in}}}\right) \\ &= \frac{1}{2}\log_2\left(1 + 2\mathrm{S_{in}} \cdot \left(\frac{2\mathrm{S_{in}}}{1 + 2\mathrm{S_{rel}} + \mathrm{S_{in}}}\right)\right). \end{aligned}$$

As $\frac{\mathrm{S_{rel}}}{\mathrm{S_{in}}} \to \infty$, $R \to \frac{1}{2}\log_2(1 + 2\mathrm{S_{in}})$, thus taking full advantage of the two independent observations $\mathbf{Y_1}$ and $\mathbf{Y_2}$. It can be seen that when $\frac{\mathrm{S_{rel}}}{\mathrm{S_{in}}}$ is large, this simple approach strictly improves upon all approaches taken in the Cover and El Gamal paper for the relay channel. The issue of how best to deal with independent observations of the source $\mathbf{X_1}$ in a distributed network is still wide open.

References

[1] R. Ahlswede and A. Kaspi. Optimal coding strategies for certain permuting channels. *IEEE IT*, 33(3):310–4, 1987.

[2] T. Cover and J. Thomas. *Elements of information theory.* Wiley, New York, first edition, 1991.

[3] E. Van der Meulen. *Transmission of information in a T-terminal discrete memoryless channel.* PhD thesis, University of California, Berkeley, 1968.

[4] E. Van der Meulen. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–54, 1971.

[5] T. Gaarder and J. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE IT*, 21(1):100–2, 1975.

[6] A. El Gamal and M. Aref. The capacity of the semideterministic relay channel. *IEEE IT*, 28(3):536, 1982.

[7] K. Kobayashi. Combinatorial structure and capacity of the permuting relay channel. *IEEE IT*, 33(6):813–26, 1987.

[8] L. Ozarow. The capacity of the white gaussian multiple access channel with feedback. *IEEE IT*, 30(4):623–29, 1984.

[9] A. Sendonaris, E. Erkip, and B. Aazhang. Increasing uplink capacity via user cooperation diversity. In *Proceedings — ISIT 1998*, Cambridge, MA, August 1998.

[10] P. Vanroose and E. Van der Meulen. Uniquely decodable codes for deterministic relay channels. *IEEE IT*, 38(4):1203–12, 1992.

[11] Z. Zhang. Partial converse for a relay channel. *IEEE IT*, 34(5):1106–10, 1988.