When is Compress-and-Forward Optimal?

Si-Hyeon Lee and Sae-Young Chung
Department of EE, KAIST, Daejeon, Korea
Email: sihyeon@kaist.ac.kr, sychung@ee.kaist.ac.kr

***Abstract*— In many known examples where compress-and-forward (CF) for relay networks is capacity achieving, it is only trivially so, i.e., it falls back to hashing without quantization. A potentially better strategy is to decode as much as possible and to compress the residual information, i.e., a combination of decode-and-forward (DF) and CF (Cover and El Gamal's Theorem 7). Indeed such a strategy was shown to be optimal by Kang and Ulukus for a certain class of diamond relay networks consisting of a source, a noisy relay, a noiseless relay, and a destination. In this paper, we discuss why it can be optimal for such channels. Furthermore, we generalize the result to a certain class of tree networks with an arbitrary number of nodes consisting of multiple cascaded diamond relay networks. We show that a combination of DF and CF is optimal for the network and its capacity is given by a simple expression. As in the diamond channel, the capacity is strictly less than the cut-set bound.**

I. Introduction

Two fundamental coding strategies for relay networks (RN) were proposed by Cover and El Gamal in [1]. In one strategy, the relay decodes the message and forwards it to the destination. This decode-and-forward (DF) coding scheme achieves the capacity for physically degraded relay channels [1]. In the other strategy, commonly called compress-and-forward (CF), the relay compresses its received block and sends the compressed information to the destination. The optimality of DF is relatively well understood, whereas it is not so clear when and how CF can be optimal in a non-trivial way.

In many examples, where CF is known to achieve capacity, it is only trivially so, i.e., no quantization is required and CF falls back to hashing. Such examples include some deterministic relay channels [2] and noisy network coding [3] applied to noiseless and interferenceless networks. If the channel from the relay to the destination is good enough such that the relay's observation can be conveyed to the destination without quantization, then CF can also be trivially optimal. The optimality of CF in such discrete cases is often translated to asymptotic optimality in additive white Gaussian noise (AWGN) channels, e.g., when the signal-to-noise ratio (SNR) of the channel between the relay and the destination tends to infinity. CF for a Gaussian RN was shown to achieve a rate within a constant number of bits from the cut-set bound in [3], [4].

One of the examples where CF is non-trivially optimal is the mod-sum relay channel studied in [5]. For this case, an optimal rate-distortion code is needed to achieve the capacity. This was generalized in [6]. Another non-trivial case is the class of diamond RN's studied in [7], which consists of a source–destination pair, one noisy relay, and one noiseless relay. For this class of diamond RN's, it was shown that a combination of CF and DF is optimal and the cut-set bound is loose [7]. In this paper, we discuss why it can be optimal, i.e., because we can always find optimal distributions such that what is compressed is a noisy observation of uncoded information. Furthermore, we generalize the result to a certain class of tree networks with an arbitrary number of nodes consisting of multiple cascaded diamond RN's. We show that a combination of DF and CF is also optimal for this class of networks and the capacity is given by a simple expression.

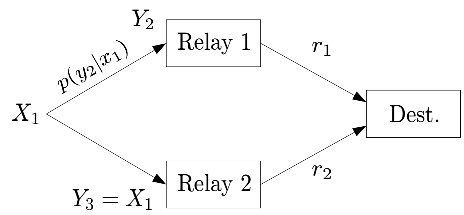

Fig. 1. Diamond RN

II. Model

A diamond RN consists of a source, a noisy relay, a noiseless relay, and a destination. See Fig. 1. The noisy and noiseless relays can send messages to the destination at rates up to r_1 and r_2 without error, respectively. The channel between the source and the noisy relay is given by $(\mathcal{X}_1, p(y_2|x_1), \mathcal{Y}_2)$ consisting of alphabets $\mathcal{X}_1, \mathcal{Y}_2$ and a conditional probability distribution $p(y_2|x_1)$, where $x_1 \in \mathcal{X}_1$ and $y_2 \in \mathcal{Y}_2$. A $\left(2^{nR}, n\right)$ code consists of a message set $\mathcal{W} = [1, 2^{nR}] \triangleq \{1, 2, ..., 2^{nR}\}$, an encoding function

$$f : \mathcal{W} \to \mathcal{X}_1^n,$$

a processing function at the noisy relay

$$h_1 : \mathcal{Y}_2^n \to [1, 2^{nr_1}],$$

a processing function at the noiseless relay

$$h_2 : \mathcal{X}_1^n \to [1, 2^{nr_2}],$$

and a decoding function

$$g : [1, 2^{nr_1}] \times [1, 2^{nr_2}] \to \mathcal{W}.$$

The source chooses an index w uniformly from the set $\mathcal{W}$ and sends $x_1^n = f(w)$. The destination decodes $\hat{w} =$

$g(h_1(y_2^n), h_2(x_1^n))$. The average probability of error for the $(2^{nR}, n)$ code is given as

$$P_e^{(n)} \triangleq \frac{1}{2^{nR}} \sum_{w \in \mathcal{W}} \Pr\left(\hat{w} \neq w | w \text{ sent}\right).$$

A rate R is said to be *achievable* if there exists a sequence of $(2^{nR}, n)$ codes such that $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity is the supremum of all achievable rates.

III. Main Result

The capacity of the diamond RN is identified as follows by Kang and Ulukus [7].

Theorem 1 (Kang and Ulukus [7]): The capacity of the diamond RN is given by (2) with cardinalities of alphabets bounded as follows:

$$|\mathcal{U}| \leq |\mathcal{X}_1| + 4 \quad (1a)$$

$$|\hat{\mathcal{Y}}_2| \leq |\mathcal{U}||\mathcal{Y}_2| + 2 \leq |\mathcal{X}_1||\mathcal{Y}_2| + 4|\mathcal{Y}_2| + 2. \quad (1b)$$

The capacity of the diamond RN is characterized as the following alternative expression whose proof is in Appendix I.

Theorem 2 (Alternative expression): The capacity of the diamond RN is given by (3) with cardinalities of alphabets bounded by (1).

Here U corresponds to the part of the message intended to be decoded by the noisy relay. The capacity characterization (3) shows that the superposition of one coded, i.e., one of $2^{n(I(U;Y_2)-\epsilon)}$ U^n's, and one uncoded, i.e., one of $2^{n(H(X_1|U)-\epsilon)}$ X^n's, codewords for $\epsilon > 0$ is always enough to achieve the capacity. After decoding U^n, the noisy relay sees a noisy version of the uncoded information. Therefore, it is not surprising to see that CF can be optimal since the uncoded part has no structure.

In the following, we also present a min-cut-like expression for the capacity of the diamond RN whose proof is in the full version of this paper.

Theorem 3 (Min-cut-like expression): The capacity of the diamond RN is given by (4) with cardinalities of alphabets bounded by (1).

We note that the relationship between the two capacity characterizations in Theorems 2 and 3 is similar to that between the two equivalent achievable rate characterizations of CF for the 3-node relay network in [1] and [8], which are given by (5) and (6), respectively, where node indices follow the convention that nodes 1, 2, and 3 are the source, relay, and destination, respectively.

The min-cut-like expression in Theorem 3 can be extended to a class of RN's with N nodes, called tree networks, in which the probability distribution has the following form:

$$p\left(y_1, ..., y_N | x_1, ..., x_N\right) = \prod_{k=1}^{N} p\left(y_k | x_{p_k}\right)$$

where p_k is the parent node of k and k is a child node of p_k. We call a node that has no parent node a root node and the node that has no child node a leaf node. We consider a tree network that has a single root node where the root node is the source, the set D of leaf nodes is the destination, and each parent node has at most one noisy child node and any number of noiseless child nodes, i.e., $y_k = x_{p_k}$ if k is a noiseless child node of p_k. Let n_k and S_k denote the noisy child node and the set of noiseless child nodes of node k, respectively and let L_k denote the subset of D that branches out from node k. For the this class of tree networks, the capacity is given as follows, whose proof is in the full version of this paper.

Theorem 4: For tree networks, the capacity is given as follows:

$$\max \min_{T} I(U_T; Y_{T^c} \setminus X_T) + I(X_T; \hat{Y}_{T^c} | U_T) - I(Y_T; \hat{Y}_T | U_T, X_T)$$

over all cuts T such that $1 \in T$, $D \subseteq T^c$, $S_k \subset T$ if $n_k \in T$, and $p_k \in T$ if $k \in T$. Here $\hat{Y}_j = X_k$ for $j \in S_k$ and $k \in [1, N]$, $Y_{T^c} \setminus X_T$ denotes the set $\{Y_j | j \in T^c, j \notin S_k \text{ for all } k \in T\}$, and the maximization is over $\prod_{k=1}^{N} p(u_k, x_k) p(y_{n_k} | x_k) p(\hat{y}_{n_k} | u_k, y_{n_k})$ with cardinalities of alphabets such that

$$|\mathcal{U}_k| \leq |\mathcal{X}_k| + 4$$

$$|\hat{\mathcal{Y}}_{n_k}| \leq |\mathcal{U}_k||\mathcal{Y}_{n_k}| + 2 \leq |\mathcal{X}_k||\mathcal{Y}_{n_k}| + 4|\mathcal{Y}_{n_k}| + 2$$

for $k \in [1, N]$.

This result is the first to show that the combination of DF and CF is capacity achieving for a non-trivial class of noisy networks with an arbitrary number of nodes.

IV. Conclusion

In this paper, we presented two equivalent capacity expressions for the diamond RN. Using the results, we showed why a combination of DF and CF can be optimal for such a network, i.e., because what is compressed is a noisy observation of uncoded information. Furthermore, we characterized the capacity of a class of noisy networks with an arbitrary number of nodes. Its proof is very different from that of noisy network coding, yet the capacity has a similar form as the noisy network coding with an additional DF part.

Acknowledgement

This work was in part supported by the IT R&D program of MKE/IITA. [2008-F-004-02] The second author would like to thank Sreeram Kannan, Young-Han Kim, and Pramod Viswanath for useful discussions.

Appendix I
Proof of Theorem 2

Let C_1 and C_2 denote the right-hand terms of (2) and (3), respectively. Let us note that the constraint on r_1 in C_1 can be easily shown to be redundant. It is trivial to show $C_2 \leq C_1$. Let us focus on $C_1 \leq C_2$. To show $C_1 \leq C_2$, it is enough to show that for $p(u, x_1)p(\hat{y}_2 | u, y_2)$ and (R, r_1, r_2) such that

$$R < I(U; Y_2) + H(X_1 | U), \quad (7a)$$

$$R = r_1 + r_2 - I(Y_2; \hat{Y}_2 | U, X_1), \quad (7b)$$

$$r_2 \geq H(X_1 | U, \hat{Y}_2), \quad (7c)$$

$$C = \max_{\substack{p(u,x_1)p(\hat{y}_2|y_2,u): \\ r_1 \geq I(Y_2;\hat{Y}_2|U,X_1) \\ r_2 \geq H(X_1|U,\hat{Y}_2)}} \min\{I(U;Y_2)+H(X_1|U), r_1+r_2-I(Y_2;\hat{Y}_2|U,X_1)\}. \tag{2}$$

$$C = \max_{\substack{p(u,x_1)p(\hat{y}_2|y_2,u): \\ r_2 \geq H(X_1|U,\hat{Y}_2) \\ r_1+r_2 \geq I(U;Y_2)+H(X_1|U)+I(Y_2;\hat{Y}_2|U,X_1)}} I(U;Y_2)+H(X_1|U) \tag{3}$$

$$C = \max_{p(u,x_1)p(\hat{y}_2|y_2,u)} \min\{I(U;Y_2)+H(X_1|U), r_2+I(U;Y_2)+I(X_1;\hat{Y}_2|U), r_1+r_2-I(Y_2;\hat{Y}_2|U,X_1)\} \tag{4}$$

$$C \geq \max_{\substack{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2): \\ I(X_2;Y_3) \geq I(Y_2;\hat{Y}_2|X_2,Y_3)}} I(X_1;\hat{Y}_2,Y_3|X_2) \tag{5}$$

$$C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} \min\{I(X_1;\hat{Y}_2,Y_3|X_2), I(X_1,X_2;Y_3)-I(Y_2;\hat{Y}_2|X_1,X_2,Y_3)\} \tag{6}$$

there exists $p(u^*,x_1^*)p(\hat{y_2}^*|u^*,y_2)$ that satisfies

$$R = I(U^*;Y_2)+H(X_1^*|U^*), \tag{8a}$$

$$R \leq r_1+r_2-I(Y_2;\hat{Y}_2^*|U^*,X_1^*), \tag{8b}$$

$$r_2 \geq H(X_1^*|U^*,\hat{Y}_2^*). \tag{8c}$$

Let $(U',X_1',\hat{Y}_2') \triangleq (X_1,X_1,\emptyset)$ where X_1 follows the marginal distribution $p(x_1)=\sum_{u\in\mathcal{U}} p(u,x_1)$. Let Q denote the random variable that has values of 1 and 2 with probability λ and $\overline{\lambda} \triangleq 1-\lambda$, respectively. Let $(U'',X_1'',\hat{Y}_2'')$ and $(U''',X_1''',\hat{Y}_2''')$ denote triplets of random variables such that

$$(U'',X_1'',\hat{Y}_2'') = \begin{cases} (U,X_1,\hat{Y}_2), & \text{for } Q=1 \\ (U',X_1',\hat{Y}_2'), & \text{for } Q=2 \end{cases}$$

$$(U''',X_1''',\hat{Y}_2''') = \begin{cases} (\emptyset,\emptyset,\emptyset), & \text{for } Q=1 \\ (U',X_1',\hat{Y}_2'), & \text{for } Q=2 \end{cases}$$

We will show the existence of $p(u^*,x_1^*)p(\hat{y_2}^*|u^*,y_2)$ that satisfies (8) for cases $I(X_1;Y_2)<R$ and $I(X_1;Y_2)\geq R$ separately. First, let us consider the case $I(X_1;Y_2)<R$. Let $U^*=(U'',Q)$, $X_1^*=X_1''$, and $\hat{Y}_2^*=(\hat{Y}_2'',Q)$. Then, we have

$$\begin{aligned} & I(U^*;Y_2)+H(X_1^*|U^*) \\ & = I(U'',Q;Y_2)+H(X_1''|U'',Q) \\ & \geq I(U'';Y_2|Q)+H(X_1''|U'',Q) \\ & = \lambda(I(U;Y_2)+H(X_1|U))+\overline{\lambda}I(X_1;Y_2) \end{aligned}$$

$$\begin{aligned} I(Y_2;\hat{Y}_2^*|U^*,X_1^*) &= I(Y_2;\hat{Y}_2''|U'',X_1'',Q) \\ &= \lambda I(Y_2;\hat{Y}_2|U,X_1) \\ &\leq I(Y_2;\hat{Y}_2|U,X_1) \end{aligned} \tag{9}$$

$$\begin{aligned} H(X_1^*|U^*,\hat{Y}_2^*) &= H(X_1''|U'',\hat{Y}_2'',Q) \\ &= \lambda H(X_1|U,\hat{Y}_2) \\ &\leq H(X_1|U,\hat{Y}_2) \end{aligned} \tag{10}$$

Since $I(U^*;Y_2)+H(X_1^*|U^*)$ becomes $I(U;Y_2)+H(X_1|U)$ and $I(X_1;Y_2)$ for $\lambda=1$ and $\lambda=0$, respectively, and it is a continuous function of λ, there exists $\lambda\in[0,1]$ such that $I(U^*;Y_2)+H(X_1^*|U^*)=R$ from the intermediate value theorem. Furthermore, (8b) and (8c) are satisfied from (9) and (10), respectively.

For the case $I(X_1;Y_2)\geq R$, let $U^*=(U''',Q)$, $X_1^*=X_1'''$, and $\hat{Y}_2^*=(\hat{Y}_2''',Q)$. Then, we get

$$\begin{aligned} & I(U^*;Y_2)+H(X_1^*|U^*) \\ & = I(U''',Q;Y_2)+H(X_1'''|U''',Q) \\ & \geq I(U''';Y_2|Q)+H(X_1'''|U''',Q) \\ & = \overline{\lambda}I(X_1;Y_2) \end{aligned}$$

$$I(Y_2;\hat{Y}_2^*|U^*,X_1^*) = I(Y_2;\hat{Y}_2'''|U''',X_1''',Q) = 0 \tag{11}$$

$$H(X_1^*|U^*,\hat{Y}_2^*) = H(X_1'''|U''',\hat{Y}_2''',Q) = 0 \tag{12}$$

Similarly as in the case $I(X_1;Y)<R$, there exists $\lambda\in[0,1]$ such that $I(U^*;Y_2)+H(X_1^*|U^*)=R$. (8b) and (8c) are satisfied from (11) and (12), respectively. ■

REFERENCES

[1] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, pp. 572–584, Sep. 1979.

[2] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1328–1329, Mar. 2008.

[3] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," in *Proc. IEEE ITW*, Cairo, Egypt, Jan. 2010, pp. 58–62.

[4] A. Avestimehr, S. N. Diggavi, and D. Tse, "Approximate capacity of Gaussian relay networks," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, Jul. 2008, pp. 474–478.

[5] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 921–930, Mar. 2009.

[6] R. Tandon and S. Ulukus, "A new upper bound on the capacity of a class of primitive relay channels," in *arXiv:cs.IT/0810.0747*, Oct. 2008.

[7] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," in *Proc. 46nd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2008.

[8] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1545–1561, Apr. 2006.