

Research Article

A New Achievable Rate and the Capacity of Some Classes of Multilevel Relay Network

Leila Ghabeli and Mohammad Reza Aref

Information Systems and Security Lab, Department of Electrical Engineering, Sharif University of Technology, P.O. Box 11365-8639 Tehran, Iran

Correspondence should be addressed to Leila Ghabeli, ghabeli@ee.sharif.edu

Received 1 August 2007; Accepted 3 March 2008

Recommended by Liang-Liang Xie

A new achievable rate based on a partial decoding scheme is proposed for the multilevel relay network. A novel application of regular encoding and backward decoding is presented to implement the proposed rate. In our scheme, the relays are arranged in feed-forward structure from the source to the destination. Each relay in the network decodes only part of the transmitted message by the previous relay. The proposed scheme differs from general parity forwarding scheme in which each relay selects some relays in the network but decodes all messages of the selected relays. It is also shown that in some cases higher rates can be achieved by the proposed scheme than previously known by Xie and Kumar. For the classes of semideterministic and orthogonal relay networks, the proposed achievable rate is shown to be the exact capacity. The application of the defined networks is very well understood in wireless networking scenarios.

Copyright © 2008 L. Ghabeli and M. R. Aref. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The relay Van der Meulen in [1], describes a single-user communication channel where a relay helps a sender-receiver pair in their communication. In [2], Cover and El Gamal proved a converse result for the relay channel, the so-called max-flow min-cut upper bound. Additionally, they established two coding approaches and three achievability results for the discrete-memoryless relay channel. They also presented the capacity of degraded, reversely degraded relay channel, and the relay channel with full feedback. In [3], partial decoding scheme or generalized block Markov encoding was defined as a special case of the proposed coding scheme by Cover and El Gamal [2, Theorem 7]. In this encoding scheme, the relay does not completely decode the transmitted message by the sender. Instead, the relay only decodes part of the message transmitted by the sender. Partial decoding scheme was used to establish the capacity of two classes of relay channels called semideterministic relay channel [3, 4] and orthogonal relay channel [5].

The last few decades have seen tremendous growth in communication networks. The most popular examples are

cellular voice, data networks, and satellite communication systems. These and other similar applications have motivated researches to extend Shannon's information theory to networks. In the case of relay networks, deterministic relay networks with no interference, first introduced by Aref [4], are named Aref networks in [6]. Aref determined the unicast capacity of such networks. The multicast capacity of Aref networks is also characterized in [6]. There also has been much interest in channels with orthogonal components, since in a practical wireless communication system, a node cannot transmit and receive at the same time or over the same frequency band. In [5], the capacity of a class of discrete-memoryless relay channels with orthogonal channels from the sender to the relay receiver and from the sender and relay to the sink is shown to be equal to the max-flow min-cut upper bound.

There also have been a lot of works that apply the proposed encoding schemes by Cover and El Gamal to the multiple relay networks [7–14]. In [7], authors generalize compress-and-forward strategy and also give an achievable rate when the relays use either decode-and-forward or compress-and-forward. Additionally, they add partial

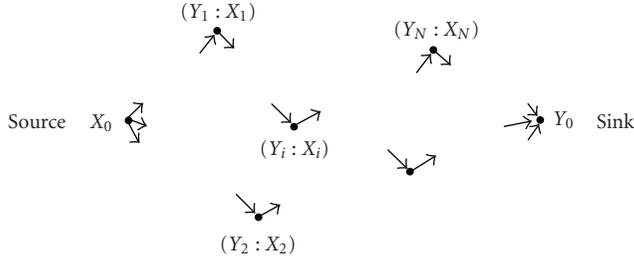


FIGURE 1: General discrete memoryless relay network [4, Figure 2.1].

decoding to the later method when there are two relays. In their scheme, the first relay uses decode-and-forward, and the second relay uses compress-and-forward. Second, relay further partially decodes the signal from first relay before compressing its observation. They made the second relay output statistically independent of the first relay and the transmitter outputs. In [8], Gupta and Kumar applied irregular encoding/successive decoding to multirelay networks in a manner similar to [4]. In [9, 10], Xie and Kumar developed regular encoding-/sliding-window decoding for multiple relays, and showed that their scheme achieves better rates than those of [4, 8]. Regular encoding/backward decoding was similarly generalized [11]. The achievable rates of the two regular encoding strategies turn out to be the same. However, the delay of sliding-window decoding is much less than that of backward decoding. Regular encoding-/sliding-window decoding is therefore currently the preferred variant of multihopping in the sense that it achieves the best rates in the simplest way. In [12, 13], parity-forwarding protocol is introduced and a structured generalization of decode-and-forward strategies for multiple-relay networks with feed-forward structure based on such protocol is proposed. In their method, each relay chooses a selective set of previous nodes in the network and decodes all messages of those nodes. Parity forwarding was shown to improve previous decode-and-forward strategies, and it achieves the capacity of new forms of degraded multirelay networks.

In [14], a generalization of partial decoding scheme was applied to multiple-relay networks and a new achievable rate was proposed. In this method, all relays in the network successively decode only part of the messages of the previous node before they arrive at the destination, In this way, using auxiliary random variables that indicate the message parts results the flexibility in defining some special classes of relay networks that the proposed rate obtain their exact capacities. For example, the capacity of feed-forward semideterministic and orthogonal relay networks that are obtained by the proposed method. To our knowledge, up to now, except the work done in [14], no other work was done for applying the partial decoding to the relay networks in which more than one relay partially decodes the message transmitted by the sender. In this paper, we generalize the results of [14] for N -relay networks and prove some theorems.

The paper is organized as follows. Section 2 introduces modeling assumptions and notations. In Section 3, some

theorems and corollaries about generalized block Markov encoding scheme or partial decoding method are reviewed. In Section 4, we introduce sequential partial decoding and drive a new achievable rate for relay networks based on this scheme. In Section 5, a class of semideterministic relay network is introduced and it is shown that the capacity of this network is obtained by the proposed method. In Section 6, we first give a review of orthogonal relay channel defined in [5], then we introduce orthogonal relay networks and obtain its capacity. Finally, some concluding remarks are provided in Section 7.

2. DEFINITIONS AND PRELIMINARIES

The discrete memoryless relay network shown in Figure 1 [4, Figure 2.1] is a model for the communication between a source X_0 and a sink Y_0 via N intermediate nodes called relays. The relays receive signals from the source and other nodes and then transmit their information to help the sink to resolve its uncertainty about the message. To specify the network, we define $2N + 2$ finite sets: $\mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_N$ and a probability transition matrix $p(y_0, y_1, \dots, y_N | x_0, x_1, \dots, x_N)$ defined for all $(y_0, y_1, \dots, y_N, x_0, x_1, \dots, x_N) \in \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_N \times \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$. In this model, X_0 is the input to the network, Y_0 is the ultimate output, Y_i is the i th relay output, and X_i is the i th relay input.

An (M, n) code for the network consists of a set of integers $\mathcal{W} = \{1, 2, \dots, M\}$, an encoding function $x_0^n : \mathcal{W} \rightarrow \mathcal{X}_0^n$, a set of relay function $\{f_{ij}\}$ such that

$$x_{ij} = f_{ij}(y_{i1}, y_{i2}, \dots, y_{i,j-1}), \quad 1 \leq i \leq N, \quad 1 \leq j \leq n, \quad (1)$$

that is, $x_{ij} \triangleq j$ th component of $x_i^n \triangleq (x_{i1}, \dots, x_{in})$, and a decoding function $g : \mathcal{Y}_0^n \rightarrow \mathcal{W}$. For generality, all functions are allowed to be stochastic functions.

Let $y_i^{j-1} = (y_{i1}, y_{i2}, \dots, y_{i,j-1})$. The input x_{ij} is allowed to depend only on the past received signals at the i th node, that is, $(y_{i1}, \dots, y_{i,j-1})$. The network is memoryless in the sense that $(y_{0i}, y_{1i}, \dots, y_{Ni})$ depends on the past $(x_0^i, x_1^i, \dots, x_N^i)$ only through the present transmitted symbols $(x_{0i}, x_{1i}, \dots, x_{Ni})$. Therefore, the joint probability mass function on $\mathcal{W} \times \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathcal{Y}_0 \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_N$ is given by

$$\begin{aligned} & p(w, x_0^n, x_1^n, \dots, x_N^n, y_0^n, y_1^n, \dots, y_N^n) \\ &= p(w) \prod_{i=1}^N p(x_{0i} | w) p(x_{1i} | y_{1i}^{i-1}) \cdots p(x_{Ni} | y_{Ni}^{i-1}) \\ & \quad \times p(y_{0i}, \dots, y_{Ni} | x_{0i}, \dots, x_{Ni}), \end{aligned} \quad (2)$$

where $p(w)$ is the probability distribution on the message $w \in \mathcal{W}$. If the message $w \in \mathcal{W}$ is sent, let $\lambda(w) \triangleq \Pr\{g(Y_0^n) \neq w | W = w\}$ denote the conditional probability of error. Define the average probability of error of the code, assuming a uniform distribution over the set of all messages $w \in \mathcal{W}$, as $\bar{P}_e^n = (1/M) \sum_w \lambda(w)$. Let $\lambda_n \triangleq \max_{w \in \mathcal{W}} \lambda(w)$ be the maximal probability of error for the (M, n) code. The rate R of an (M, n) code is defined to be $R = (1/n) \log M$

bits/transmission. The rate R is said to be achievable by the network if, for any $\epsilon > 0$, and for all n sufficiently large, there exists an (M, n) code with $M \geq 2^{nR}$ such that $\bar{P}_e^n < \epsilon$. The capacity \mathcal{C} of the network is the supremum of the set of achievable rates.

3. GENERALIZED BLOCK MARKOV ENCODING

In [3], generalized block Markov encoding is defined as a special case of [2, Theorem 7]. In this encoding scheme, the relay does not completely decode the transmitted message by the sender. Instead the relay only decodes part of the message transmitted by the sender. A block Markov encoding timeframe is again used in this scheme such that the relay decodes part of the message transmitted in the previous block and cooperates with the sender to transmit the decoded part of the message to the sink in current block. The following theorem expresses the obtained rate via generalized block Markov encoding.

Theorem 1 (see [3]). *For any relay network $(\mathcal{X}_0 \times \mathcal{X}_1, p(y_0, y_1 | x_0, x_1), \mathcal{Y}_0 \times \mathcal{Y}_1)$, the capacity \mathcal{C} is lower-bounded by*

$$\mathcal{C} \geq \max_{p(x_0, x_1)} \min \{I(X_0 X_1; Y_0), I(U; Y_1 | X_1) + I(X_0; Y_0 | X_1 U)\}, \quad (3)$$

where the maximum is taken over all joint probability mass functions of the form

$$p(u, x_0, x_1, y_0, y_1) = p(u, x_0, x_1) \cdot p(y_0, y_1 | x_0, x_1) \quad (4)$$

such that $U \rightarrow (X_0, X_1) \rightarrow (Y_0, Y_1)$ form a Markov chain.

If we choose the random variable $U = X_0$, it satisfies the Markovity criterion and the result of block Markov coding directly follows as

$$\mathcal{C} \geq \max_{p(x_0, x_1)} \min \{I(X_0 X_1; Y_0), I(X_0; Y_1 | X_1)\}. \quad (5)$$

The above expression introduces the capacity of degraded relay channel as shown in [2]. Moreover, by substituting $U = Y_1$ in (3), the capacity of semideterministic relay channel in which y_1 is a deterministic function of x_0 and x_1 .

Corollary 1. *If y_1 is a deterministic function of x_0 and x_1 , then*

$$\mathcal{C} \geq \max_{p(x_0, x_1)} \min \{I(X_0 X_1; Y_0), H(Y_1 | X_1) + I(X_0; Y_0 | X_1 Y_1)\}. \quad (6)$$

In the next section, we apply the concept of Theorem 1 to the relay networks with N relays and prove the main theorem of this paper.

4. SEQUENTIAL PARTIAL DECODING

In this section, we introduce sequential partial decoding method and drive a new achievable rate for N -relay networks. In sequential partial decoding, the message of the sender is divided into N parts. The first part is directly

decoded by the sink, while the other parts are decoded by the first relay. With the same way, at each relay, one part of the message is directly decoded by the sink, while the other parts are decoded by the next relay. In the next blocks, the sender and the relays cooperate with each other to remove the uncertainty of the sink about the individual parts of the messages.

Sequential partial decoding scheme is useful in the cases that the relays are located in feed-forward structure from the sender to the sink with at most distance with each other in such a way that each node is able to decode some parts of the message of the previous node, while the sink is sensitive enough to be able to directly decode the remaining parts of the messages of the sender and the relays. The rate obtained by this method is expressed in the following theorem.

Theorem 2. *For any relay network $(\mathcal{X}_0 \times \mathcal{X}_1 \times \dots \times \mathcal{X}_N, p(y_0, y_1, \dots, y_N | x_0, x_1, \dots, x_N), \mathcal{Y}_0 \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_N)$, the capacity \mathcal{C} is lower-bounded by*

$$\begin{aligned} \mathcal{C} \geq \sup \min \{ & I(X_0, X_1, \dots, X_N; Y_0), \\ & \min_{1 \leq i \leq N} \{ I(U_i; Y_i | X_i \{X_l U_l\}_{l=i+1}^N) \\ & + I(\{X_l\}_{l=0}^{i-1}; Y_0 | \{X_l U_l\}_{l=i}^N) \} \}, \end{aligned} \quad (7)$$

where the supremum is over all joint probability mass functions $p(u_1, \dots, u_N, x_0, x_1, \dots, x_N)$ on

$$\mathcal{U}_1 \times \dots \times \mathcal{U}_N \times \mathcal{X}_0 \times \dots \times \mathcal{X}_N \quad (8)$$

such that

$$(U_1, \dots, U_N) \rightarrow (X_0, \dots, X_N) \rightarrow (Y_0, \dots, Y_N) \quad (9)$$

form a Markov chain.

Proof. In this encoding scheme, the source message is split into $(N + 1)$ parts, $w_{N0}, w_{N-1,0}, \dots, w_{00}$. The first relay decodes messages w_{N0}, \dots, w_{10} ; the second relay decodes w_{N0}, \dots, w_{20} ; and so on the N th relay decodes only w_{N0} . Each relay retransmits its decoded messages to the sink using the same codebook as the source, that is, regular encoding is used. Backward decoding is used at all nodes to decode messages, starting from the last block and going backward to the first block.

We consider B blocks of transmission, each of n symbols. A sequence of $B - N$ messages,

$$\begin{aligned} w_{00,i} \times w_{10,i} \times \dots \times w_{N0,i} \in & [1, 2^{nR_0}] \times [1, 2^{nR_1}] \\ & \times \dots \times [1, 2^{nR_N}], \quad i = 1, 2, \dots, B - N, \end{aligned} \quad (10)$$

will be sent over the channel in nB transmissions. In each n -block $b = 1, 2, \dots, B$, we will use the same set of codewords. We consider only the probability of error in each block as the total average probability of error can be upper-bounded by the sum of the decoding error probabilities at each step, under the assumption that no error propagation from the previous steps has occurred [15].

Random coding

The random codewords to be used in each block are generated as follows.

(1) Choose 2^{nR_N} i.i.d. x_N^n each with probability $p(x_N^n) = \prod_{i=1}^n p(x_{Ni})$. Label these as $x_N^n(w_{NN})$, $w_{NN} \in [1, 2^{nR_N}]$.

(2) For every $x_N^n(w_{NN})$, generate 2^{nR_N} i.i.d. u_N^n with probability

$$p(u_N^n | x_N^n(w_{NN})) = \prod_{i=1}^n p(u_{N,i} | x_{N,i}(w_{NN})). \quad (11)$$

Label these $u_N^n(w_{NN}, w_{N,N-1})$, $w_{N,N-1} \in [1, 2^{nR_N}]$.

(3) For each $(u_N^n(w_{NN}, w_{N,N-1}), x_N^n(w_{NN}))$, generate $2^{nR_{N-1}}$ i.i.d. x_{N-1}^n each with probability

$$p(x_{N-1}^n | u_N^n(w_{NN}, w_{N,N-1}), x_N^n(w_{NN})) = \prod_{i=1}^n p(x_{N-1,i} | u_{N,i}(w_{NN}, w_{N,N-1}), x_{N,i}(w_{NN})). \quad (12)$$

Label these $x_{N-1}^n(w_{NN}, w_{N,N-1}, w_{N-1,N-1})$, $w_{N-1,N-1} \in [1, 2^{nR_{N-1}}]$.

For every $l \in \{N-1, \dots, 1\}$, with the same manner as previous, we generate u_l^n and x_{l-1}^n in the steps $a = 2N - 2(l-1)$ and $b = 2N - 2(l-1) + 1$, respectively, as follows.

(a) For each

$$u_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right), \quad l+1 \leq j \leq N, \quad (13)$$

$$x_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}} \right), \quad l \leq j \leq N,$$

generate 2^{nR_l} i.i.d. u_l^n with probability

$$p \left(u_l^n \mid \left\{ u_j^n \left(\{w_{km}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, j\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right) \right\}_{j=l+1}^N, \left\{ x_j^n \left(\{w_{km}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, j-1\}} \right) \right\}_{j=l}^N \right) = \prod_{i=1}^n p \left(u_{l,i} \mid \left\{ u_{j,i} \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right) \right\}_{j=l+1}^N, \left\{ x_{j,i} \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}} \right) \right\}_{j=l}^N \right). \quad (14)$$

Label these $u_l^n(\{w_{km}\}_{m \in \{N, \dots, l\}, k \in \{N, \dots, m\}}, \{w_{k,l-1}\}_{k \in \{N, \dots, l\}})$, $\{w_{k,l-1} \in [1, 2^{nR_k}]\}_{k \in \{N, \dots, l\}}$.

(b) For each

$$x_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}} \right), \quad l \leq j \leq N,$$

$$u_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right), \quad l \leq j \leq N, \quad (15)$$

generate $2^{nR_{l-1}}$ i.i.d. x_{l-1}^n with probability

$$p \left(x_{l-1}^n \mid \left\{ u_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right) \right\}_{j=l}^N, \left\{ x_j^n \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}} \right) \right\}_{j=l}^N \right) = \prod_{i=1}^n p \left(x_{l-1,i} \mid \left\{ u_{j,i} \left(\{w_{km}\}_{m \in \{N, \dots, j-1\}, k \in \{N, \dots, m\}}, \{w_{k,j-1}\}_{k \in \{N, \dots, j\}} \right) \right\}_{j=l}^N, \left\{ x_{j,i} \left(\{w_{km}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}} \right) \right\}_{j=l}^N \right). \quad (16)$$

Label these $x_{l-1}^n(\{w_{km}\}_{m \in \{N, \dots, l-1\}, k \in \{N, \dots, m\}})$, $w_{l-1,l-1} \in [1, 2^{nR_{l-1}}]$.

In the above random coding strategy, m and k denote the relay number and the message part number, respectively. The m th relay decodes $\{w_{km}\}_{k \in \{N, \dots, m\}}$ and the $(m-1)$ th relay decodes $\{w_{k,m-1}\}_{k \in \{N, \dots, m-1\}}$, where at the m th relay for each message part $k \in \{N, \dots, m\}$, the index w_{km} represents the index $w_{k,m-1}$ of the previous block. In this coding construction, $N(N+1)$ indices are used in total.

As an example, for two-relay network, the transmitter and the relay encoders send the following codewords:

$$x_0^n(1, 1, 1, w_{20,h}, w_{10,h}, w_{00,h}), \quad x_1^n(1, 1, 1), \quad x_2^n(1); \quad (17)$$

in block $h = 1$, the following codewords:

$$x_0^n(1, w_{20,h-1}, w_{10,h-1}, w_{20,h}, w_{10,h}, w_{00,h}), \quad x_1^n(1, \hat{w}_{20,h-1}, \hat{w}_{10,h-1}), \quad x_2^n(1); \quad (18)$$

in each block $h = 2$, the following codewords:

$$x_0^n(w_{20,h-2}, w_{20,h-1}, w_{10,h-1}, w_{20,h}, w_{10,h}, w_{00,h}), \quad x_1^n(\hat{w}_{20,h-2}, \hat{w}_{20,h-1}, \hat{w}_{10,h-1}), \quad x_2^n(\hat{w}_{20,h-2}); \quad (19)$$

in each block $h = 3, \dots, B-2$, the following codewords

$$x_0^n(w_{20,B-3}, w_{20,B-2}, w_{10,B-2}, 1, 1, 1), \quad x_1^n(\hat{w}_{20,B-3}, \hat{w}_{20,B-2}, \hat{w}_{10,B-2}), \quad x_2^n(\hat{w}_{20,B-3}); \quad (20)$$

in block $h = B-1$, and the following codewords:

$$x_0^n(w_{20,B-2}, 1, 1, 1, 1, 1), \quad x_1^n(\hat{w}_{20,B-2}, 1, 1), \quad x_2^n(\hat{w}_{20,B-2}); \quad (21)$$

in block $h = B$. Figure 2 shows the individual parts of the messages that should be decoded by the relays and the sink. It can be inferred from Figure 2 that $w_{11,h} = w_{10,h-1}$, $w_{22,h} = w_{21,h-1}$, $w_{21,h} = w_{20,h-1}$ or $w_{22,h} = w_{20,h-2}$.

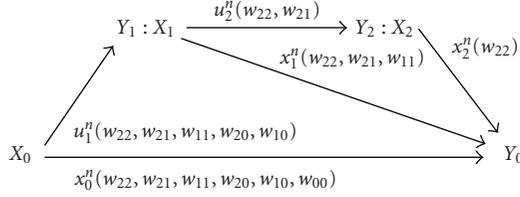


FIGURE 2: Schematic diagram of sequential partial decoding for two-relay network.

Decoding

Assume that at the end of block $(h - 1)$, the i th relay knows $\{w_{ki,i+1}, w_{ki,i+2}, \dots, w_{ki,h-1}\}_{k \in \{N, \dots, i\}}$ or equivalently $\{w_{k0,1}, w_{k0,2}, \dots, w_{k0,h-i-1}\}_{k \in \{N, \dots, i\}}$. At the end of block h , decoding is performed in the following manner.

Decoding at the relays

By knowing $\{w_{k0,1}, w_{k0,2}, \dots, w_{k0,h-i-1}\}_{k \in \{N, \dots, i\}}$, the i th relay determines $\{w_{ki,h} = w_{k0,h-i}\}_{k \in \{N, \dots, i\}}$ such that

$$\left(\begin{array}{l} u_i^n(\{\hat{w}_{km,h}\}_{m \in \{N, \dots, i\}, k \in \{N, \dots, m\}}, \{\hat{w}_{k,i-1,h}\}_{k \in \{N, \dots, i\}}), \\ y_i^n(h) \left\{ u_i^n(\{\hat{w}_{km,h}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}, \right. \\ \left. \{\hat{w}_{k,j-1,h}\}_{k \in \{N, \dots, j\}}) \right\}_{l=i+1}^N, \\ \left. \left\{ x_i^n(\{\hat{w}_{km,h}\}_{m \in \{N, \dots, j\}, k \in \{N, \dots, m\}}) \right\}_{l=i}^N \right) \in A_\epsilon^n, \end{array} \quad (22)$$

$\{\hat{w}_{ki,h} = w_{ki,h}\}_{k \in \{N, \dots, i\}}$, or similarly $\{\hat{w}_{k0,h-i} = w_{k0,h-i}\}_{k \in \{N, \dots, i\}}$ with high probability if

$$\sum_{k=i}^N R_k < I(U_i; Y_i | \{U_l\}_{l=i+1}^N \{X_l\}_{l=i}^N) \quad (23)$$

and n is sufficiently large.

Decoding at the sink

Decoding at the sink is performed in backward manner in $N + 1$ steps until all $\{w_{k0,h-N}\}_{k \in \{N, \dots, 1\}, h \in \{B, \dots, N+1\}}$ are decoded by the sink.

(1) Decoding $\{w_{N0}\}$

In block B , the sink determines the unique $\hat{w}_{NN,B} = \hat{w}_{N0,B-N}$ such that

$$(u_N^n(\hat{w}_{NN,B}, 1), x_N^n(\hat{w}_{NN,B}), y_0^n(B)) \in A_\epsilon^n \quad (24)$$

or equivalently,

$$(u_N^n(\hat{w}_{N0,B-N}, 1), x_N^n(\hat{w}_{N0,B-N}), y_0^n(B)) \in A_\epsilon^n, \quad (25)$$

$\hat{w}_{N0,B-N} = w_{N0,B-N}$ with high probability if

$$R_N < I(X_N U_N; Y_0) \quad (26)$$

and n is sufficiently large. By knowing $\hat{w}_{N0,B-N}$, in block $B - 1$, the sink determines the unique $\hat{w}_{NN,B-1} = \hat{w}_{N0,B-N-1}$ such that

$$(u_N^n(\hat{w}_{N0,B-N-1}, \hat{w}_{N0,B-N}), x_N^n(\hat{w}_{N0,B-N-1}), y_0^n(B-1)) \in A_\epsilon^n, \quad (27)$$

$\hat{w}_{N0,B-N-1} = w_{N0,B-N-1}$ with high probability if (26) is satisfied and n is sufficiently large. This way continues until first block such that all $\{w_{N0,h-N}\}_{h \in \{B, \dots, N+1\}}$ are decoded by the sink.

(2) Decoding $\{w_{N-1,0}\}$

By knowing $\{w_{N0,h-N}\}_{h \in \{B, \dots, N+1\}}$, in block $B - 1$, the sink determines the unique $\hat{w}_{N-1,N-1,B-1} = \hat{w}_{N-1,0,B-N}$ such that

$$\left(\begin{array}{l} u_{N-1}^n(\hat{w}_{NN,B-1}, \hat{w}_{N,N-1,B-1}, \hat{w}_{N-1,N-1,B-1}, 1, 1), \\ x_{N-1}^n(\hat{w}_{NN,B-1}, \hat{w}_{N,N-1,B-1}, \hat{w}_{N-1,N-1,B-1}), \\ u_N^n(\hat{w}_{NN,B-1}, \hat{w}_{N,N-1,B-1}), \\ \left. x_N^n(\hat{w}_{NN,B-1}), \quad y_0^n(B-1) \right) \in A_\epsilon^n \end{array} \right) \quad (28)$$

or equivalently,

$$\left(\begin{array}{l} u_{N-1}^n(\hat{w}_{N0,B-N-1}, \hat{w}_{N0,B-N}, \hat{w}_{N-1,0,B-N}, 1, 1), \\ x_{N-1}^n(\hat{w}_{N0,B-N-1}, \hat{w}_{N0,B-N}, \hat{w}_{N-1,0,B-N}), \\ u_N^n(\hat{w}_{N0,B-N-1}, \hat{w}_{N0,B-N}), \\ \left. x_N^n(\hat{w}_{N0,B-N-1}), \quad y_0^n(B-1) \right) \in A_\epsilon^n, \end{array} \right) \quad (29)$$

$\hat{w}_{N-1,0,B-N} = w_{N-1,0,B-N}$ with high probability if

$$R_{N-1} < I(X_{N-1} U_{N-1}; Y_0 | X_N U_N) \quad (30)$$

and n is sufficiently large. By knowing $\hat{w}_{N-1,0,B-N}$, in block $B - 2$, the sink determines the unique $\hat{w}_{N-1,N-1,B-2} = \hat{w}_{N0,B-N-1}$ such that

$$\left(\begin{array}{l} u_{N-1}^n(\hat{w}_{N0,B-N-2}, \hat{w}_{N0,B-N-1}, \hat{w}_{N-1,0,B-N-1}, \\ \hat{w}_{N0,B-N}, \hat{w}_{N-1,0,B-N}), \\ x_{N-1}^n(\hat{w}_{N0,B-N-2}, \hat{w}_{N0,B-N-1}, \hat{w}_{N-1,0,B-N-1}), \\ u_N^n(\hat{w}_{N0,B-N-2}, \hat{w}_{N0,B-N-1}), \\ \left. x_N^n(\hat{w}_{N0,B-N-2}), y_0^n(B-2) \right) \in A_\epsilon^n, \end{array} \right) \quad (31)$$

$\hat{w}_{N-1,0,B-N-1} = w_{N-1,0,B-N-1}$ with high probability if (30) is satisfied and n is sufficiently large. This way continues until first block such that all $\{w_{N-1,0,h-N}\}_{h \in \{B, \dots, N+1\}}$ are decoded by the sink.

(3) Decoding $\{w_{i0}\}$

By knowing $\{w_{k0,h-N}\}_{h \in \{B, \dots, N+1\}, k \in \{N, \dots, i+1\}}$, in block $B+i-N$, the sink determines the unique $\hat{w}_{ii,B-N+i} = \hat{w}_{i0,B-N}$ such that

$$\begin{pmatrix} \{u_l^n(\{\hat{w}_{km,B-N+i}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, l\}}), \\ \{\hat{w}_{k,l-1,B-N+i}\}_{k \in \{N, \dots, l\}}\}, \\ x_l^n(\{\hat{w}_{km,B-N+i}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, l-1\}})\}_{l=i}^N \\ y^n(B-N+i), \end{pmatrix} \in A_\epsilon^n \quad (32)$$

or equivalently,

$$\begin{pmatrix} \{u_l^n(\{\hat{w}_{k0,B-N+i-m}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, l\}}), \\ \{\hat{w}_{k0,B-N+i-l+1}\}_{k \in \{N, \dots, l\}}\}, \\ x_l^n(\{\hat{w}_{k0,B-N+i-m}\}_{k \in \{N, \dots, m\}, m \in \{N, \dots, l-1\}})\}_{l=i}^N \\ y^n(B-N+i), \end{pmatrix} \in A_\epsilon^n, \quad (33)$$

$\hat{w}_{i0,B-N} = w_{i0,B-N}$ with high probability if

$$R_i < I(X_i U_i; Y_0 | \{X_l U_l\}_{l=i+1}^N) \quad (34)$$

and n is sufficiently large.

This way continues until first block such that all $\{w_{i0,h-N}\}_{h \in \{B, \dots, N+1\}}$ are decoded by the sink.

By knowing $U_0 = 0$, (34) reduces to the following constraint for $i = 0$;

$$R_0 < I(X_0; Y_0 | \{X_l U_l\}_{l=1}^N). \quad (35)$$

Now, For each $1 \leq i \leq N$, we have

$$\begin{aligned} R_{ti} &= \sum_{k=0}^N R_k \\ &= R_0 + \sum_{k=1}^{i-1} R_k + \sum_{k=i}^N R_k \\ &\stackrel{(a)}{<} I(X_0; Y_0 | \{X_l U_l\}_{l=1}^N) + \sum_{k=1}^{i-1} R_k + I(U_i; Y_i | X_i \{X_l U_l\}_{l=i+1}^N) \\ &\stackrel{(b)}{<} I(X_0; Y_0 | \{X_l U_l\}_{l=1}^N) + \sum_{k=1}^{i-1} I(X_k U_k; Y_0 | \{X_l U_l\}_{l=k+1}^N) \\ &\quad + I(U_i; Y_i | X_i \{X_l U_l\}_{l=i+1}^N) \\ &\stackrel{(c)}{=} I(\{X_l\}_{l=0}^{i-1}; Y_0 | \{X_l U_l\}_{l=i}^N) + I(U_i; Y_i | X_i \{X_l U_l\}_{l=i+1}^N), \end{aligned} \quad (36)$$

where (a) follows from (23) and (35). (b) follows from (34). (c) follows from chain rule for information and (9). For $i = 0$, by respect to the fact that $U_{N+1} = 0$, we have

$$R_{t0} < I(X_0, X_1, \dots, X_N; Y_0), \quad (37)$$

equations (36) and (37) along with $R = \min_{0 \leq i \leq N} R_{ti}$ result in (7).

This completes the proof. \square

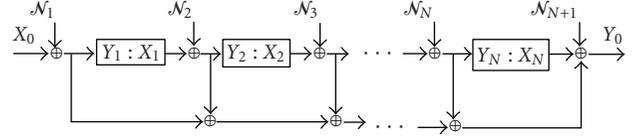


FIGURE 3: A degraded chain network with additive noises \mathcal{N}_k , $1 \leq k \leq N$, [13, Figure 5].

Remarks

(1) By putting $U_l = X_{l-1}$ for $1 \leq l \leq N$ in (7), that means omitting partial decoding and assuming that each relay decodes all messages of the previous relay, the following rate is the result:

$$\begin{aligned} \mathcal{C} &\geq \sup_{p(x_0, x_1, \dots, x_N)} \min \{I(X_0, X_1, \dots, X_N; Y_0), \\ &\quad \min_{1 \leq i \leq N} \{I(X_{i-1}; Y_i | \{X_l\}_{l=i}^N) \\ &\quad \quad + I(\{X_l\}_{l=0}^{i-2}; Y_0 | \{X_l\}_{l=i-1}^N)\}\}. \end{aligned} \quad (38)$$

In [13], the above rate is obtained as a special case of parity-forwarding method in which each relay selects the message of the previous relay, and it is stated that (38) is the capacity of degraded chain network as shown in Figure 3. This point can be regarded as a special example that two schemes coincide.

(2) By comparing (38) with the rate proposed by Xie and Kumar in [10, Theorem 3-1], as

$$\mathcal{C} \geq \sup_{p(x_0, x_1, \dots, x_N)} \min_{1 \leq i \leq N} \{I(\{X_l\}_{l=0}^{i-1}; Y_i | \{X_l\}_{l=i}^N)\}; \quad (39)$$

it is seen that in the cases which the following relation

$$I(\{X_l\}_{l=0}^{i-2}; Y_0 | \{X_l\}_{l=i-1}^N) > I(\{X_l\}_{l=0}^{i-1}; Y_i | \{X_l\}_{l=i-1}^N) \quad (40)$$

is true for $2 \leq i \leq N$, (38) yields higher rates than (39).

(3) In our proposed rate (7), we offer more flexibility than parity forwarding scheme [13], by introducing auxiliary random variables that indicate partial parts of the messages. By this priority, we can achieve the capacity of some other forms of relay networks such as semideterministic and orthogonal relay networks as shown in the next section. However, our scheme is limited by the assumption that each relay only decodes part of the message transmitted by the previous relay.

5. A CLASS OF SEMIDETERMINISTIC RELAY NETWORKS

seminet A class of semideterministic multirelay networks In this section, we introduce a class of semideterministic relay network and show that the capacity of such network is obtained by using the proposed method, that is, it coincides with the max-flow min-cut upper bound. Consider the semideterministic relay networks with N relays as shown in

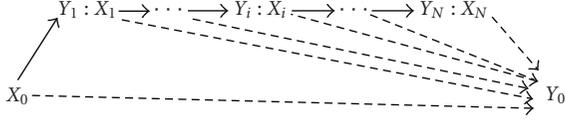


FIGURE 4: A class of semideterministic multirelay networks.

Figure 4 in which $y_k = h_k(x_{k-1}, \dots, x_N)$ for $k = 1, \dots, N$, are deterministic functions. In this figure, deterministic and nondeterministic links are shown by solid and dash lines, respectively. It can be easily proved that the capacity of this network is obtained by the proposed method and it coincides with max-flow min-cut upper bound. It is expressed in the following theorem.

Theorem 3. For a class of semideterministic relay network $(\mathcal{X}_0 \times \mathcal{X}_1 \times \dots \times \mathcal{X}_N, p(y_0, y_1, \dots, y_N | x_0, x_1, \dots, x_N), \mathcal{Y}_0 \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_N)$ having $y_k = h_k(x_{k-1}, \dots, x_N)$ for $k = 1, \dots, N$, the capacity \mathcal{C} is given by

$$\mathcal{C} \geq \sup_{p(x_0, x_1, \dots, x_N)} \min \left\{ I(X_0, X_1, \dots, X_N; Y_0), \min_{1 \leq i \leq N} H(Y_0 Y_i | \{X_k\}_{k=i}^N) - H(Y_0 | \{X_k\}_{k=0}^N) \right\}. \quad (41)$$

Proof. The achievability is proved by replacing $U_k = Y_k$ for $k = 1, \dots, N$, in (7). The converse follows immediately from the max-flow min-cut theorem for general multiple-node networks stated in [16, Theorem 15.10.1], where the node set is chosen to be $\{0\}, \{0, 1\}, \dots, \{0, 1, \dots, N\}$ sequentially, and with the following equation:

$$\begin{aligned} & I(X_0, \dots, X_{i-1}; Y_i, \dots, Y_N, Y_0 | X_i, \dots, X_N) \\ &= H(Y_i, \dots, Y_N, Y_0 | X_i, \dots, X_N) \\ &\quad - H(Y_i, \dots, Y_N, Y_0 | X_0, \dots, X_N) \\ &= H(Y_i, Y_0 | X_i, \dots, X_N) - H(Y_0 | X_0, \dots, X_N). \quad \square \end{aligned} \quad (42)$$

6. A CLASS OF ORTHOGONAL RELAY NETWORKS

In this section, we introduce a class of orthogonal relay networks that is a generalization of orthogonal relay channel [5]. First, we define orthogonal relay channel.

A relay channel with orthogonal components is a relay channel where the channel from the transmitter to the relay is orthogonal to the channel from the sender and relay to the sink. In other words, transmission on direct channel from the sender to the sink does not affect the reception at the relay and also transmission at the channel from the sender to the relay does not affect the received signal at the sink. This channel is defined in as follows [5].

Definition 1. A discrete-memoryless relay channel is said to have orthogonal components if the sender alphabet $\mathcal{X}_0 = \mathcal{X}_D \times \mathcal{X}_R$ and the channel can be expressed as

$$p(y_0, y_1 | x_0, x_1) = p(y_0 | x_D, x_1) p(y_1 | x_R, x_1) \quad (43)$$

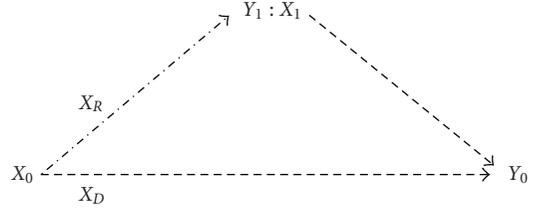


FIGURE 5: A class of orthogonal relay channel.

for all

$$(x_D, x_R, x_1, y_0, y_1) \in \mathcal{X}_D \times \mathcal{X}_R \times \mathcal{X}_1 \times \mathcal{Y}_0 \times \mathcal{Y}_1. \quad (44)$$

The class A relay channel is illustrated in Figure 5, where the channels in the same frequency band are shown by the lines with the same type. The capacity is given by the following theorem.

Theorem 4 (see [5, Theorem]). The capacity of the relay channel with orthogonal components is given by

$$\mathcal{C} = \max \min \{ I(X_D, X_1; Y_0), I(X_R; Y_1 | X_1) + I(X_D; Y_0 | X_1) \}, \quad (45)$$

where the maximum is taken over all joint probability mass functions of the form

$$p(x_1, x_D, x_R) = p(x_1) p(x_D | x_1) p(x_R | x_1). \quad (46)$$

Generalized block Markov coding is used for the proof of achievability part by assuming joint probability mass function of the form (46). The converse part of the theorem is proved for all joint probability mass function $p(x_1, x_D, x_R)$ only based on the orthogonality assumption (48) or equivalently the following Markov chains:

$$X_D \longrightarrow (X_1, X_R) \longrightarrow Y_1, \quad (X_R, Y_1) \longrightarrow (X_1, X_D) \longrightarrow Y. \quad (47)$$

Now, we introduce a class of relay networks with orthogonal components where the channels reach at each node uses the same frequency band while the channels diverge from each node uses different frequency bands. By this assumption, the network with N relays (intermediate nodes) uses $(N + 1)$ frequency bands. The network is defined as follows.

Definition 2. A discrete-memoryless relay networks with N relays is said to have orthogonal components if the sender and the relays alphabet $\mathcal{X}_k = \mathcal{X}_{kR} \times \mathcal{X}_{kD}$, for $k = 0, \dots, N-1$, and the channel can be expressed as

$$\begin{aligned} & p(y_0, \dots, y_N | x_0, \dots, x_N) \\ &= p(y_0 | \{x_{kD}\}_{k=0}^{N-1}, x_N) \prod_{i=1}^N p(y_i | \{x_{kD}\}_{k=i}^{N-1}, \{x_{kR}\}_{k=i-1}^{N-1}, x_N) \end{aligned} \quad (48)$$

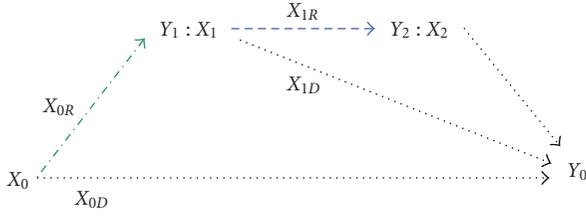


FIGURE 6: A class of orthogonal relay networks.

or equivalently,

$$\left(\{Y_k\}_{k=1}^N, \{X_{kR}\}_{k=0}^{N-1} \right) \rightarrow \left(\{X_{kD}\}_{k=0}^{N-1}, X_N \right) \rightarrow Y_0, \quad (49)$$

$$\begin{aligned} \left(\{Y_k\}_{k=1}^{i-1}, \{X_{kD}\}_{k=0}^{i-1}, \{X_{kR}\}_{k=0}^{i-2} \right) &\rightarrow \left(\{X_{kD}\}_{k=i}^{N-1}, \{X_{kR}\}_{k=i-1}^{N-1}, X_N \right) \\ &\rightarrow Y_i, \quad 1 \leq i \leq N. \end{aligned} \quad (50)$$

Orthogonal relay networks with $N = 2$ are illustrated in Figure 6, where the channels in the same frequency band are shown by the same line type. The channel from the sender to the first relay is shown by a dash-dot line. The channel between two relays is shown by a dash line. The channels converging at the sink are shown by dot lines. The capacity for the networks of Figure 6 is given by the following theorem.

Theorem 5. For the network depicted in Figure 6, the capacity is given by

$$\mathcal{C} = \sup \min \left\{ \begin{array}{l} I(X_{0D}, X_{1D}, \dots, X_{N-1,D}, X_N; Y_0), \\ \min_{1 \leq i \leq N} \left\{ I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \right. \\ \left. + I(\{X_{lD}\}_{l=0}^{i-1}; Y_0 \mid \{X_{lD}\}_{l=i}^{N-1} X_N) \right\} \end{array} \right\}. \quad (51)$$

Proof. The achievability is proved By replacing $X_k = (X_{kD}, X_{kR})$, for $k = 1, \dots, N-1$ and $U_k = X_{k-1,R}$ for $k = 1, \dots, N$, in (7) and assuming joint probability mass function of the form

$$\prod_{i=0}^{N-1} \left\{ p(x_{iD} \mid \{x_{lD}\}_{l=i+1}^{N-1}, x_N) p(x_{iR} \mid \{x_{lD}x_{lR}\}_{l=i+1}^{N-1}, x_N) \right\} p(x_N) \quad (52)$$

as follows:

$$\begin{aligned} I(X_0, \dots, X_N; Y_0) &= I(\{X_{lD}X_{lR}\}_{l=0}^{N-1} X_N; Y_0) \\ &= I(\{X_{lD}\}_{l=0}^{N-1}, X_N; Y_0) + I(\{X_{lR}\}_{l=0}^{N-1}; Y_0 \mid \{X_{lD}\}_{l=0}^{N-1}, X_N) \\ &\stackrel{(a)}{=} I(\{X_{lD}\}_{l=0}^{N-1}, X_N; Y_0); \end{aligned} \quad (53)$$

$$\begin{aligned} &I(U_i; Y_i \mid X_i \{X_l U_l\}_{l=i+1}^N) + I(\{X_l\}_{l=0}^{i-1}; Y_0 \mid \{X_l U_l\}_{l=i}^N) \\ &= I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lD}X_{lR}\}_{l=0}^{i-1}; Y_0 \mid X_{i-1,R} \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &= I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lD}\}_{l=0}^{i-1} \{X_{lR}\}_{l=0}^{i-2}; Y_0 \mid X_{i-1,R} \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &= I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lD}\}_{l=0}^{i-1}; Y_0 \mid X_{i-1,R} \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lR}\}_{l=0}^{i-2}; Y_0 \mid X_{i-1,R} \{X_{lD}\}_{l=0}^{N-1} \{X_{lR}\}_{l=i}^{N-1} X_N) \\ &\stackrel{(b)}{=} I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lD}\}_{l=0}^{i-1}; Y_0 \mid X_{i-1,R} \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\stackrel{(c)}{=} I(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1} X_N) \\ &\quad + I(\{X_{lD}\}_{l=0}^{i-1}; Y_0 \mid \{X_{lD}\}_{l=i}^{N-1} X_N), \end{aligned} \quad (54)$$

where (a) and (b) follow from (49). (c) follows from the fact that according to (52), we have

$$\{X_{lR}\}_{l=i-1}^{N-1} \rightarrow (\{X_{lD}\}_{l=i}^{N-1} X_N) \rightarrow \{X_{lD}\}_{l=0}^{i-1}, \quad (55)$$

this along with (50) results in

$$\{X_{lR}\}_{l=i-1}^{N-1} \rightarrow (\{X_{lD}\}_{l=i}^{N-1} X_N) \rightarrow Y_0. \quad (56)$$

The converse follows immediately from the max-flow min-cut theorem for general multiple-node networks stated in [16, Theorem 15.10.1], where the node set is chosen to be $\{0\}, \{0, 1\}, \dots, \{0, 1, \dots, N\}$ sequentially, and with the following equations:

$$\begin{aligned} &I(X_0, \dots, X_{i-1}, Y_i, \dots, Y_N, Y_0 \mid X_i, \dots, X_N) \\ &= I(\{X_k\}_{k=0}^{i-1}; Y_i, \dots, Y_N, Y_0 \mid \{X_k\}_{k=i}^{N-1}, X_N) \\ &= I(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_i, \dots, Y_N, Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N) \\ &= I(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N) \\ &\quad + \sum_{l=i+1}^N I(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_l \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^{l-1}) \\ &\quad + I(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^N) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} I\left(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + I\left(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^N\right) \\
&= H\left(Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) - H\left(Y_i \mid \{X_{kD}X_{kR}\}_{k=0}^{N-1}, X_N\right) \\
&\quad + I\left(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^N\right) \\
&\stackrel{(b)}{=} H\left(Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad - H\left(Y_i \mid X_{i-1,R}, \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + I\left(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^N\right) \\
&= I\left(X_{i-1,R}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + I\left(\{X_{kD}X_{kR}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_k\}_{k=i}^N\right) \\
&= I\left(X_{i-1,R}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + H\left(Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_l\}_{l=i}^N\right) \\
&\quad - H\left(Y_0 \mid \{X_{kD}X_{kR}\}_{k=0}^{N-1}, X_N, \{Y_l\}_{l=i}^N\right) \\
&\stackrel{(c)}{=} I\left(X_{i-1,R}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + H\left(Y_0 \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N, \{Y_l\}_{l=i}^N\right) \\
&\quad - H\left(Y_0 \mid \{X_{kD}\}_{k=0}^{N-1}, X_N\right) \\
&\stackrel{(d)}{<} I\left(X_{i-1,R}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + H\left(Y_0 \mid \{X_{kD}\}_{k=i}^{N-1}, X_N\right) - H\left(Y_0 \mid \{X_{kD}\}_{k=0}^{N-1}, X_N\right) \\
&= I\left(X_{i-1,R}; Y_i \mid \{X_{kD}X_{kR}\}_{k=i}^{N-1}, X_N\right) \\
&\quad + I\left(\{X_{kD}\}_{k=0}^{i-1}; Y_0 \mid \{X_{kD}\}_{k=i}^N, X_N\right), \tag{57}
\end{aligned}$$

where (a) and (b) follow from (50). (c) follows from (49). (d) follows from the fact that conditioning reduces entropy. For the set $\{0, 1, \dots, N\}$, according to (53), the first term of (51) are obtained.

We have shown that

$$\mathcal{C} = \sup \min \left\{ \begin{array}{l} I(X_{0D}, X_{1D}, \dots, X_{N-1,D}, X_N; Y_0), \\ \min_{1 \leq i \leq N} \left\{ I\left(X_{i-1,R}; Y_i \mid \{X_{lD}X_{lR}\}_{l=i}^{N-1}, X_N\right) \right. \\ \left. + I\left(\{X_{lD}\}_{l=0}^{i-1}; Y_0 \mid \{X_{lD}\}_{l=i}^{N-1}, X_N\right) \right\} \end{array} \right\}. \tag{58}$$

The maximization in (51) is over the choice of joint probability mass function $\{p(x_N) \prod_{i=0}^{N-1} p(x_{iD}, x_{iR})\}$. Without loss of generality, we can restrict the joint probability mass functions to be of the form (52).

This completes the proof of the theorem. \square

7. CONCLUSION

This paper presents a new achievable rate based on a partial decoding scheme for the multilevel relay network. A novel application of regular encoding and backward decoding is presented to implement the proposed rate. In the proposed scheme, the relays are arranged in feed-forward structure from the source to the destination. Each relay in the network decodes only part of the transmitted message by the previous relay. The priorities and differences between the proposed method with similar previously known methods such as general parity forwarding scheme and the proposed rate by Xie and Kumar are specified. For the classes of semideterministic and orthogonal relay networks, the proposed achievable rate is shown to be the exact capacity. One of the applications of the defined networks is in wireless networks which have nondeterministic or orthogonal channels.

ACKNOWLEDGMENTS

This work was supported by Iranian National Science Foundation (INSF) under Contract no. 84,5193-2006. Some parts of this paper were presented at the IEEE International Symposium on Information Theory, Nice, France, June 2007.

REFERENCES

- [1] E. C. Van der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, no. 1, pp. 120–154, 1971.
- [2] T. M. Cover and A. EL Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.
- [3] A. EL Gamal and M. R. Aref, "The capacity of the semideterministic relay channel," *IEEE Transactions on Information Theory*, vol. 28, no. 3, p. 536, 1982.
- [4] M. R. Aref, "Information flow in relay networks," Ph.D. dissertation, Stanford University, Stanford, Calif, USA, October 1980.
- [5] A. EL Gamal and S. Zahedi, "Capacity of relay channels with orthogonal components," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1815–1817, 2005.
- [6] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2425–2432, 2006.
- [7] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.
- [8] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1877–1894, 2003.
- [9] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 748–767, 2004.
- [10] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1348–1358, 2005.

- [11] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proceedings of the 41st Annual Allerton Conference on Communications, Control, and Computing*, pp. 1074–1083, Monticello, Ill, USA, October 2003.
- [12] P. Razaghi and W. Yu, "Parity forwarding for multiple-relay networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 1678–1682, Seattle, Wash, USA, July 2006.
- [13] P. Razaghi and W. Yu, "A structured generalization of decode-and-forward for the multiple relay network," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 271–275, Nice, France, June 2007.
- [14] L. Ghabeli and M. R. Aref, "A new achievable rate and the capacity of a class of semideterministic relay networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 281–285, Nice, France, June 2007.
- [15] C.-M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1160–1165, 1989.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2006.