Capacity of a Modulo-Sum Simple Relay Network

Youvaraj T. Sagar, Hyuck M. Kwon, and Yanwu Ding
Electrical Engineering and Computer Science, Wichita State University
Wichita, Kansas 67260, USA, {ytsagar; hyuck.kwon; yanwu.ding}@wichita.edu

Abstract

This paper presents the capacity of a modulo-sum simple relay network. In previous work related to this paper, capacity was characterized for the case where the noise was transmitted to the relay. And the closed-form capacity was derived only for the noise with a Bernoulli-(1/2) distribution. However, in this paper, the source is transmitted to the relay, and a more general case of noise with an arbitrary Bernoulli-(p) distribution, p ∈ [0,0.5], is considered. The relay observes a corrupted version of the source, uses a quantize-and-forward strategy, and transmits the encoded codeword through a separate dedicated channel to the destination. The destination receives both from the relay and source. This paper assumes that the channel is discrete and memoryless. After deriving the achievable capacity theorem (i.e., the forward theorem) for the binary symmetric simple relay network, this paper proves that the capacity is strictly below the cut-set bound. In addition, this paper presents the proof of the converse theorem. Finally, this paper extends the capacity of the binary symmetric simple relay network to that of an m-ary modulo-sum relay network.

Index Terms – Channel capacity; relay network; modulo-sum channel; quantize-and-forward; single-input single-output; cut-set bound.

---

This work was partly sponsored by the Army Research Office under DEPSCoR ARO Grant W911NF-08-1-0256, and by NASA under EPSCoR CAN Grant NNX08AV84A.

08/30/2009
I. INTRODUCTION

The relay network is a channel that has one sender and one receiver, with a number of intermediate nodes acting as relays to assist with the communication between sender and receiver. This paper exchanges the terminology of the relay channel in [1] with the relay network frequently because here a network is defined as consisting of more than two nodes [2], whereas a channel is for communication between two nodes. The simplest relay network or channel has one sender, one receiver, and one relay node. Fig. 1 shows this type of relay network, which is called a “simple” relay network.

The first original model of a relay network was introduced by van der Meulen in 1971 [3]. After that, extensive research was done to find the upper bounds, cut-set bounds, and exact capacity for this network. In 1979, Thomas and Gamal obtained the capacity for a special class of channels called physically degraded relay channels [4]. In that paper, they discussed the capacity of the relay channel with feedback and found an upper bound for a simple relay network, which is shown in Fig. 1. Later, Gamal found the capacity for a special class of relay channels called “semideterministic relay channels,” which he discussed in a paper [5]. Then, Kim found the capacity for a class of deterministic relay channels in [6], where he modeled the simple relay network as a noiseless channel between the relay and the destination. Also, van der Meulen corrected his previous upper bound on the capacity of the simple relay network with and without delay in a paper [7].

Using Kim’s results in [6], Aleksic et al. modeled the channel between the relay and the destination as a modular sum noise channel in [8]. Binary side information or channel state information is transmitted to the relay in [8]. He mentioned that the capacity of the simple relay network is not yet known. Recently, Tandon found a new upper bound for the simple relay network with general noise, obtained the capacity for symmetric binary erasure relay channel, and compared them with the cut-set bound in [9].

Aleksic et al. in [8] introduced a corrupting variable to the noiseless channel in [6], whereby the noise in the direct channel between the source and the destination is transmitted to the relay. The relay observes a corrupted version of the noise and has a separate dedicated channel to the destination. For this case, the capacity was characterized in [8]. However, the closed-form capacity was derived only for the noise with a Bernoulli-$(p = 1/2)$ distribution.

The objective of this paper is to find the capacity of the simple modular sum relay network and show that its capacity is strictly below the cut-set bound [4]. This paper also presents a closed-form capacity for a general case, such as for any $p$ where the source is transmitted to both the relay and the destination.

This paper considers all noisy channels, i.e., from the source to the destination, from the source to the relay, and from the relay to the destination, as shown in Fig. 1, where all noisy channels are binary symmetric channels (BSCs) with a certain crossover probability, e.g., $p$. This paper also derives the capacity for this class of relay channels. In other words, the capacity of a modulo-sum simple relay network is presented here. The capacity proof for the binary symmetric simple relay network and the proof for the converse depend crucially on the input distribution.
For the BSC, a uniform input distribution at the source is assumed because this distribution maximizes the entropy of the output (or the capacity) regardless of additive noise. Furthermore, because of the uniform input distribution, the output of a binary symmetric relay network is independent of additive noise. After presenting the proof for the capacity of a binary symmetric simple relay network, this paper proves that the capacity obtained is strictly below the cut-set bound by using the results in [4]. Finally, this paper shows the converse theorem for this class of networks.

Section II describes the system model and presents the capacity of the binary symmetric simple relay network. Proofs of the converse and achievability for this theorem are provided in appendices A and B. Section III discusses the cut-set bound for the binary symmetric simple relay network and presents the numerical analysis results. Section IV extends the capacity to the $m$-ary modular additive case. Finally, Section V concludes the paper.

II. SYSTEM MODEL AND NETWORK CAPACITY

Fig. 2 shows a realistic binary phase-shift keying (BPSK) system under additive white Gaussian noise (AWGN), where $X$ and $Y$ are the binary input and output signal, respectively. Here, $Y$ is obtained with a hard decision on the demodulated signal.

Fig. 3 shows a BSC with the crossover probability $p$ equivalent to the realistic communication system, as shown in Fig. 2. Here, the crossover probability $p$ is equal to $Q\left(\sqrt{2E_b/N_0}\right)$, where $Q(x)$ is the Gaussian tail probability function. 

08/30/2009
\( Q(\alpha) = \int_{-\infty}^{\infty} (1/\sqrt{2\pi}) e^{-t^2/2} \, dt \), and \( E_b \) and \( N_0 \) denote the bit energy and the one-side AWGN power spectral density, respectively.

This paper models a channel between any adjacent nodes in Fig. 1 as a BSC that has one sender, one receiver (or destination), and one relay node [1]. The random variable \( Y \) represents the received signal through the direct channel and is written as \( Y = X \oplus Z \), where \( X \) and \( Z \) denote the transmitted and noise random variable with distribution \( \text{Ber}(1/2) \) and \( \text{Ber}(p) \), respectively, and \( \oplus \) denotes the binary modulo-sum, i.e., \( Z = 1 \) with probability \( p \), and \( Z = 0 \) with probability \( (1 - p) \).

The simple relay network in Fig. 1 can be redrawn as Fig. 4. Here, the relay node has an input \( Y_1 \) and an output \( X_1 \). The relay node observes the corrupted version of \( X \), i.e., \( Y_1 = X \oplus N_1 \), encodes it using a codebook \( \mathcal{U}^n \) of jointly typical strong sequences [1], and transmits the code symbol \( X_1 \) through another separate BSC to the destination node, where \( \mathcal{U} \), \( n \), and \( N_1 \) denote the alphabet of code symbols, the codeword length, and the noise random variable at the relay with distribution \( \text{Ber}(\delta) \), respectively. The destination receives both \( Y \), through the direct channel, and \( S_0 = X_1 \oplus N_2 \), through the relay node, where \( N_2 \sim \text{Ber}(\epsilon) \) represents the noise at the destination for the relay network.

Note that the binary modulo-sum and the BSC can be extended to an \( m \)-ary modulo-sum and an \( m \)-ary symmetric channel (MSC).

To the authors’ knowledge, there is no network capacity expression in the literature, even for the simple relay network shown in Fig. 4. Only the capacity of a deterministic relay channel, i.e., the case of \( N_2 = 0 \) in Fig. 4, is presented in [6]. The capacity of a relay network by replacing \( X \) with \( Z \), i.e., the case where the relay observes a corrupted version of the direct channel noise \( Z \), is presented in [8]. This paper presents the capacity of the simple relay network shown in Fig. 4 in the following theorem.

**Theorem 1**: The capacity \( C \) of the binary symmetric simple relay network shown in Fig. 4 is

\[
C = \max_{p(u|y_1):U \subseteq \mathcal{U}} \left\{ 1 + H(Y|U) - H(Z) - H(X|U) \right\},
\]

where the maximization is over the \( U \)’s conditional probability density function (p.d.f.) given \( Y_1 \); the cardinality of the alphabet \( \mathcal{U} \), is bounded by \( |\mathcal{U}| \leq |\mathcal{Y}| + 2 \); and \( R_0 \) is the capacity for the channel between \( X_1 \) and \( S_0 \), which can be written as

\[
R_0 = \max_{p(x_1)} I(X_1; S_0).
\]

The closed-form network capacity for the simple relay network shown in Fig. 4 can be written as

\[
C = 1 + \mathcal{H}([\epsilon \ast \delta] \ast p) - \mathcal{H}(p) - \mathcal{H}(\epsilon \ast \delta).
\]

Here, \( H(X) \) and \( I(X;Y) \) are the entropy of \( X \) and the mutual information between \( X \) and \( Y \), respectively [1]; \( \mathcal{H}(\alpha) \) is the binary entropy function written as \( \mathcal{H}(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha) \); and \( \alpha \ast \beta = \alpha (1 - \beta) + (1 - \alpha) \beta \) [10].

**Proof**: Formal proofs for the achievability and the converse of Theorem 1 are presented in appendix A and B, respectively.

08/30/2009
Note that if the direct channel noise $Z$ is transmitted through the relay rather than $X$, then (1) becomes (3) of [8] or (4), written as

$$C = \max_{p(u|y_1):I(U;Y_1) \leq R_0} 1 - H(Z|U). \quad (4)$$

This is because $H(Y|U)$ and $H(Z)$ in (1) will become 1 and $H(X) = 1$, respectively.

### III. CUTSET BOUND AND ANALYTICAL RESULTS

This section proves that the capacity of the binary symmetric simple relay network in Fig. 4 is strictly below the cut-set bound, except for the two trivial points at $R_0 = 0$ and $R_0 = 1$. The capacity in (1) can be upper-bounded by the cut-set bound as

$$C \leq \max_{p(x,x_1)} \min \{I(X,X_1;Y), I(X;Y,Y_1)\} \quad (5)$$

where the Ford-Fulkerson theorem [11], [4] is applied to the simple relay network in Fig. 4. Using (5), Theorem 2 can be established.

**Theorem 2:** The cut-set bound for the capacity of the binary simple relay network shown in Fig. 4 can be written as

$$C \leq \min \{1 - H(Z) + R_0, 1 - H(Z) + 1 - H(N_1)\}$$

$$= \min \{1 - \mathcal{H}(p) + R_0, 1 - \mathcal{H}(p) + 1 - \mathcal{H}(\delta)\}. \quad (6)$$

**Proof:** A formal proof for Theorem 2 is presented in appendix C.

Figs. 5(a) and 5(b) show the capacity in bits per transmission versus $R_0$ bits for $\delta = 0.1$, when $p = 0.1$ and $p = 0.5$, respectively. If $p = 0.5$, then the results are the same as those in [8]. Only the closed form of the capacity for the special case of $p = 0.5$ was analyzed and presented in [8], where the capacity $C$ of the binary simple relay network was obtained by replacing $X$ with $Z$ at the relay input shown in Fig. 4 and written as [8]

$$C = 1 - \mathcal{H}(\mathcal{H}^{-1}\{1 - R_0\} * \delta). \quad (7)$$

Here $\mathcal{H}^{-1}(\cdot)$ is the inverse of $\mathcal{H}(p)$ in the domain $p \in [0,0.5]$. Note that the capacity in (3) of this paper is valid for a general $p$ between 0 and 0.5, whereas the one in (34) of [8] or (7) is valid for only $p = 0.5$.

Note that the capacity in (3) is strictly below the cut-set bound in (6). Refer to Figure 5(b).

### IV. CAPACITY FOR M-ARY MODULO-SUM RELAY NETWORK

This section extends the capacity derived for the binary symmetric simple relay network to the $m$-ary modular additive relay network. The received signal at the destination node can be written as $Y = X + Z \mod m$. The relay observes the corrupted version of $X$, i.e., $Y_1 = X + N_1 \mod m$, and the relay also has a separate channel to the destination: $S_0 = X_1 + N_2 \mod m$ with a capacity $R_0 = \max_{p(x_1)} I(X_1;S_0)$. Therefore, (1) becomes (8) in Theorem 3.

08/30/2009
**Theorem 3**: The capacity $C$ of the symmetric $m$-ary modulo-sum simple relay network is

$$C = \max_{p(u|y_1):|(U,Y)|\leq R_0} \left\{ m + H(Y|U) - H(Z) - H(X|U) \right\}$$

(8)

where maximization is over the conditional $U$’s p.d.f. given $Y_1$ with $|U| \leq |Y_1| + 2$, and $R_0$ is defined in (2).

**Proof**: The achievability for Theorem 3 follows the same steps as Theorem 1 by changing the binary to the $m$-ary case. Also, the uniform input distribution at the source maximizes the entropy of the output, regardless of the additive noise. Furthermore, because of the uniform input distribution, the output of an $m$-ary modulo-sum relay network is independent of the additive noise. Therefore, (8) holds true. The converse for Theorem 3 also holds true using the same steps of Theorem 1 by changing the binary modulo-sum to the $m$-ary modulo-sum.

![Fig. 5(a). Capacity of a binary symmetric simple relay network shown in Fig. 4 for $\delta = 0.1$ and $p = 0.1$.](image)

![Fig. 5(b). Capacity of a binary symmetric simple relay network shown in Fig. 4 for $\delta = 0.1$ and $p = 0.5$.](image)

V. CONCLUSIONS

It has been an open problem to find the capacity of the simple relay network. This paper presented the closed form capacity of the binary symmetric simple relay network. Also, this paper extended the capacity for the binary to the $m$-ary modulo-sum symmetric simple relay network. Two conditions are necessary for the derivation of this capacity: (1) a uniform Bernoulli-(1/2) input distribution, and (2) a modular additive channel between the two adjacent nodes. Using these conditions, both proofs for the achievability and the converse of the capacity theorem were presented. Furthermore, this paper derived the cut-set bound and presented the numerical results for this network. Finally, this paper determined that the capacity is strictly below the cut-set bound and achievable using a quantize-and-forward strategy at the relay.
This paper uses the same steps as in [8] and modifies them for Fig. 4. This appendix proves the achievability for Theorem 1 as follows:

1. **Generation of Codebook.** With the best input distribution of $X \sim Ber(1/2)$, $2^{nR}$ number of codewords $X^n(w)$ are independently and identically generated, indexed with $w \in \{1, 2, \ldots \}$, and used at the source. Similarly, with the best $p(u|y_1)$ distribution satisfying $I(U; Y_1) \leq R_0$, $2^{nI(U; Y_1)}$ number of codewords $U^n(v)$ are independently and identically generated, indexed with $v \in \{1, 2, \ldots \}$, and used at the relay.

2. **Encoding.** The generated two codebooks $X^n$ and $U^n$ are revealed to (source and destination) and (relay and destination), respectively. Now, to send the $i$th block of the source, the channel encoder at the source sends codeword $X^n(w_i)$ with index $w_i$ to both the destination and the relay node. The relay finds the jointly strongly typical sequence $U^n(v_i)$ from its codebook $U^n$, using the previously observed $Y^n_t(i - 1)$ word, which is a corrupted version of $X^n(w_{i-1})$. And the relay encodes the index $v_i$, and transmits it to the destination.

3. **Decoding.** The destination, upon receiving the two sequence information at different time slots from the source and the relay, decodes separately. After decoding the received words for index $v_i$, the destination looks for the jointly strongly typical sequence $X^n(w_{i-1})$ with $U^n(v_i)$ and $Y^n(i - 1)$.

4. **Analysis of Probability of Error.** The decoding of the received word $Y^n(w_i)$ and $S^n_t(v_i)$ are done at different time slots, i.e., when $X^n(w_i)$ is transmitted, the destination receives $Y^n(w_i)$, but not from the relay. For the next time slot, the relay transmits index $v_i$ to the destination, and the transmitter sends the codeword $X^n(w_{i+1})$ to the destination simultaneously. Now, the destination for the second time slot receives the two vectors $Y^n(w_{i+1})$ and $S^n(v_i)$. Therefore, for the first block transmission, the destination receives only from the direct channel. And for the last block, no message is transmitted from the source but rather only from the relay to the destination. Here, the last block from the source can be neglected for the rate calculation as the number of blocks $\rightarrow \infty$.

For the error probability analysis, this paper considers the average error probability over all codewords generated. Since the codebook construction is symmetric, the average error probability does not depend on a particular index sent. Therefore, it is assumed that $X^n(1)$ was transmitted. Then, this paper applies a jointly strongly typical decoding strategy. An error event can happen when either the triple $(X^n(1), Y^n(w_i), U^n(v_i))$ is not jointly strongly typical or the received words $(Y^n(w_i), U^n(v_i))$ are jointly strongly typical with an incorrect codeword $X^n(w_i \neq 1)$. The probability of mapping a correct codeword into an incorrect codeword is $2^{-nI(X;Y,U)}$ [1, pp. 327]. This event is considered
an “undetected codeword error event,” denoted by $E_i$. The probability of event $E_i$ is given as [1, Lemma 10.6.2]

$$P(E_i) < 2^{nR_2 - n[I(X;Y,U) - \epsilon_1]}.$$  \hspace{1cm} (9)

If $n$ is sufficiently large and $R < I(X;Y,U) - \epsilon_1$, then the average error probability can be made arbitrarily small, i.e., $P(E) < \epsilon$. Hence, it requires

$$R < I(X;Y,U).$$  \hspace{1cm} (10)

Since $X \sim Ber(1/2)$, the mutual information is already maximized over $p(x)$. From the chain rule [1, p.34],

$$I(X;Y,U) = I(X;Y|U) + I(X;U).$$  \hspace{1cm} (11)

This can be rewritten as

$$I(X;Y,U) = H(Y|U) - H(Y|X,U) + H(X) - H(X|U).$$  \hspace{1cm} (12)

Even if $X \sim Ber(1/2)$, it can be shown that $Y$ and $Y_1$ are not independent, and hence $Y$ and $U$ are dependent. Therefore, $H(Y|U)$ in (12) cannot be simplified to $H(Y) = 1$, as done in [8]. Equation (12) can be rewritten as

$$I(X;Y,U) = H(Y|U) - H(Y|X,U) + 1 - H(X|U)$$  \hspace{1cm} (13)

$$= H(Y|U) - H(X + Z|X,U) + 1 - H(X|U)$$  \hspace{1cm} (14)

$$= H(Y|U) - H(Z|X,U) + 1 - H(X|U)$$  \hspace{1cm} (15)

$$(a) \quad 1 + H(Y|U) - H(Z|X) - H(X|U)$$  \hspace{1cm} (16)

$$(b) \quad 1 + H(Y|U) - H(Z) - H(X|U)$$  \hspace{1cm} (17)

where

(a) follows from the fact that $Z$ and $U$ are independent; and

(b) follows from the fact that $Z$ and $X$ are independent.

Hence,

$$C = \max_{p(u|y_1):I(U;Y_1) \leq R_0} \{1 + H(Y|U) - H(Z) - H(X|U)\}. \hspace{1cm} (18)$$

The capacity can be expressed in a closed-form through which it is simple to calculate the capacity for given input and noise distributions. The capacity can now be evaluated as

$$C = 1 + \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) * \delta) * p) - H(Z) - \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) * \delta). \hspace{1cm} (19)$$

Proof of (19) is presented as follows: The mutual information between $U$ and $Y_1$ in Fig. 4 is given as

$$I(U;Y_1) = H(Y_1) - H(Y_1|U).$$  \hspace{1cm} (20)

Applying the constraint $I(U;Y_1) \leq R_0$ in (18) to (20), it can be written as

$$H(Y_1|U) \geq H(Y_1) - R_0.$$  \hspace{1cm} (21)

From Corollary 4 in [10] called “Mrs. Gerber’s Lemma,” if $H(X_0|U) \geq \nu$, then $H(Y_0|U) \geq \mathcal{H}(p_0 * \mathcal{H}^{-1}(\nu))$, where $X_0$ and $Y_0$ are the binary random input and output random variables of a BSC with crossover probability $p_0$, respectively [9]. Equality holds if and only if code symbols $X_0$ are independent of each other. Now, this Corollary 4 is applied for the case of input $Y_1$ and output $X$ in Fig. 4 as

08/30/2009
\[ X = Y_1 + N_1 \pmod{2} = Y_1 \oplus N_1. \] (22)

Since \( X \sim \text{Ber}(1/2) \) and \( N_1 \sim \text{Ber}(\delta) \), \( Y_1 \) and \( N_1 \) are independent, if

\[ H(Y_1|U) \geq \alpha, \] (23)

then

\[ H(X|U) \geq \mathcal{H}(\mathcal{H}^{-1}(\alpha) \ast \delta). \] (24)

Equality in (24) holds if and only if \( Y_1|U \sim \text{Ber}(\beta = \mathcal{H}^{-1}(\alpha)) \). Here, \( \alpha \triangleq H(Y_1) - R_0 \) from (21).

And the standard rate-distortion theory in [8, p. 925] is applied, which says that “\( Y_1|U \sim \text{Ber}(\mathcal{H}^{-1}(H(Y_1) - R_0)) \), is precisely the \( U \) that minimizes the Hamming distortion of \( Y_1 \) under a rate constraint \( R_0 \).” With this, the minimum achievable average distortion \( \beta \) under rate constraint \( R_0 \) must satisfy \( \mathcal{H}(\beta) = H(Y_1|U) = H(Y_1) - R_0 \), and hence \( Y_1|U \) must be \( \text{Ber}(\beta) \) [8, p. 925]. Also since \( X \sim \text{Ber}(1/2), Y_1 \) is \( \text{Ber}(1/2) \). Therefore, the optimal distribution of \( U \) is also \( \text{Ber}(1/2) \). Hence, \( H(Y_1) = 1 \) and \( \alpha = 1 - R_0 \). By substituting \( \alpha \) into (24), it can be written for the equality case as

\[ H(X|U) = \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) \ast \delta). \] (25)

By substituting (25) into (1), (1) can be rewritten as

\[ C = 1 + H(Y|U) - H(Z) - \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) \ast \delta). \] (26)

Again, “Mrs. Gerber’s Lemma” is applied for \( H(Y|U) \) in (26). In other words, if \( H(X|U) \geq \nu \), then \( H(Y|U) \geq \mathcal{H}(p \ast \mathcal{H}^{-1}(\nu)) \) with equality, if and only if \( X \) given \( U \) is a \( \text{Ber}(\mathcal{H}^{-1}(\nu)) \), i.e., \( H(X|U) = \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) \ast \delta) \). Therefore,

\[ H(Y|U) = \mathcal{H}(p \ast \{\mathcal{H}^{-1}(1 - R_0) \ast \delta\}) \] (27)

By substituting (27) into (26), (26) can be rewritten as

\[ C = 1 + \mathcal{H}(p \ast \{\mathcal{H}^{-1}(1 - R_0) \ast \delta\}) - H(Z) - \mathcal{H}(\mathcal{H}^{-1}(1 - R_0) \ast \delta). \] (28)

Equation (28) can be simplified further as follows: From (2), \( R_0 \) is the capacity of a BSC between \( X_1 \) and \( S_0 \) with crossover probability \( \varepsilon \). Using the uniform distribution for input \( X_1 \), (2) can be rewritten as

\[ R_0 = 1 - \mathcal{H}(\varepsilon) \] (29)

where \( N_2 \sim \text{Ber}(\varepsilon) \). Now, by substituting (29) into (28), (28) can be rewritten as

\[ C = 1 + \mathcal{H}((\mathcal{H}^{-1}(1 - [1 - \mathcal{H}(\varepsilon)]) \ast \delta) \ast p) - H(Z) - \mathcal{H}(\mathcal{H}^{-1}(1 - [1 - \mathcal{H}(\varepsilon)]) \ast \delta) \]
\[ = 1 + \mathcal{H}((\varepsilon \ast \delta) \ast p) - H(Z) - \mathcal{H}(\mathcal{H}^{-1}(\mathcal{H}(\varepsilon)) \ast \delta) \]
\[ = 1 + \mathcal{H}((\varepsilon \ast \delta) \ast p) - H(p) - \mathcal{H}(\varepsilon \ast \delta) \] (30)

Therefore, this completes the proof of the achievability for Theorem 1.
APPENDIX B

(Proof: Converse [Reverse] for Theorem 1)

The proof for the converse is close to that in [1, Theorem 15.8.1] or [8, Theorem 1]. However, a detailed proof is provided here again because Fig. 4 in this current paper has different characteristics from those in Fig. 15.32 of [1] or Fig. 2 of [8]. First, this paper derives Lemma 1.

Lemma 1: Let \((X^n, Z^n, N_1^n, N_2^n)\) be independently and identically distributed (i.i.d.) random variables of each other, with distributions of \(X \sim Ber(1/2)\), \(Z \sim Ber(p)\), \(N_1 \sim Ber(\delta)\), and \(N_2 \sim Ber(\epsilon)\), as shown in Fig. 4. And let the input at the relay and the received word at the destination be written as \(Y_1^n = X^n + N_1^n\) and \(S_0^n = X^n + N_2^n\), respectively. Then, the following inequality holds for any encoding scheme applied at the relay with the constraint \(I(U; Y_1) \leq R_0\):

\[
H(X^n|S_0^n) \geq nH(X|U) \quad (33)
\]

where \(U = (U_Q, Q)\) is an auxiliary random vector satisfying \(|U| \leq |Y_1| + 2\), and \(Q\) is a time sharing random variable with i.i.d. over \(\{1, 2, \ldots, n\}\).

Proof for Lemma 1: The proof for Lemma 1 follows the same steps in [8, Lemma 1]. To prove Lemma 1 in this current paper for any encoding scheme with the constraint, it is necessary to show

\[
H(X^n|S_0^n) \geq nH(X|U) \quad (33)
\]

where \(U = (U_Q, Q)\) is an auxiliary random vector satisfying \(|U| \leq |Y_1| + 2\), and \(Q\) is a time sharing random variable with i.i.d. over \(\{1, 2, \ldots, n\}\).

From the chain rule, the left part of (33) can be written as

\[
H(X^n|S_0^n) = \sum_{i=1}^{n} H(X_i|S_0^n, X_{1}, X_{2}, \ldots, X_{i-1}) \quad (34)
\]

where

(a) follows from the definition of \(X_{i-1} \triangleq X_1, X_2, \ldots, X_{i-1}\);

(b) follows from the conditional entropy property, i.e., \(H(X) \geq H(X|Y)\); and

(c) follows from the fact that \(X_{i-1} \rightarrow (S_0^n, Y_1^{i-1}) \rightarrow X_i\) forms a Markov chain because \(X_i\) and \(X_{i-1}\) are independent, and hence, \(X_{i-1}\) and \(X_i\) are conditionally independent for any given encoding scheme \(U_i \triangleq (S_0^n, Y_1^{i-1})\). Hence, \(X_i\) is only affected by \(X_{i-1}\) through \((S_0^n, Y_1^{i-1})\).

Then, from (37),

\[
H(X^n|S_0^n) \geq \sum_{i=1}^{n} H(X_i|U_i) \quad (38)
\]

From Fig. 4, \(X^n \rightarrow Y_1^n \rightarrow X_1^n \rightarrow S_0^n\) forms a Markov chain. Using the data processing inequality,

\[
I(X^n_i; S_0^n) \geq I(Y_1^n; S_0^n) \quad (39)
\]

where

(a) follows from the definition of \(X_{i-1} \triangleq X_1, X_2, \ldots, X_{i-1}\);

(b) follows from the conditional entropy property, i.e., \(H(X) \geq H(X|Y)\); and

(c) follows from the fact that \(X_{i-1} \rightarrow (S_0^n, Y_1^{i-1}) \rightarrow X_i\) forms a Markov chain because \(X_i\) and \(X_{i-1}\) are independent, and hence, \(X_{i-1}\) and \(X_i\) are conditionally independent for any given encoding scheme \(U_i \triangleq (S_0^n, Y_1^{i-1})\). Hence, \(X_i\) is only affected by \(X_{i-1}\) through \((S_0^n, Y_1^{i-1})\).

Then, from (37),

\[
H(X^n|S_0^n) \geq \sum_{i=1}^{n} H(X_i|U_i) \quad (38)
\]

From Fig. 4, \(X^n \rightarrow Y_1^n \rightarrow X_1^n \rightarrow S_0^n\) forms a Markov chain. Using the data processing inequality,

\[
I(X^n_i; S_0^n) \geq I(Y_1^n; S_0^n) \quad (39)
\]

where

(a) follows from the definition of \(X_{i-1} \triangleq X_1, X_2, \ldots, X_{i-1}\);

(b) follows from the conditional entropy property, i.e., \(H(X) \geq H(X|Y)\); and

(c) follows from the fact that \(X_{i-1} \rightarrow (S_0^n, Y_1^{i-1}) \rightarrow X_i\) forms a Markov chain because \(X_i\) and \(X_{i-1}\) are independent, and hence, \(X_{i-1}\) and \(X_i\) are conditionally independent for any given encoding scheme \(U_i \triangleq (S_0^n, Y_1^{i-1})\). Hence, \(X_i\) is only affected by \(X_{i-1}\) through \((S_0^n, Y_1^{i-1})\).
\[ (c) \sum_{i=1}^{n} I(Y_{1i}; S_{0}^{n}, Y_{i}^{i-1}) \]
\[ \geq \sum_{i=1}^{n} I(Y_{1i}; U_{i}) \]  

where

(a) and (b) follow from the chain rule and from the definition of \( Y_{1i}^{i-1} \equiv Y_{11}, Y_{12}, \ldots, Y_{1i-1} \); and

(c) follows from the fact that \( Y_{1i} \) and \( Y_{1i}^{i-1} \) are independent, i.e., \( I(Y_{1i}; Y_{1i}^{i-1}) = 0 \) because \( X \sim Ber(1/2) \); hence, \( Y_{1i} \sim Ber(1/2) \) too, and \( (Y_{1i}, N_{i}) \) is an independent pair, regardless of the noise distribution \( N_{i} \sim Ber(\delta) \).

The lemma in [1, §7.9.2] says that the capacity of a discrete memoryless channel is lower-bounded by
\[ I(X_{1}^{n}; S_{0}^{n}) \geq \sum_{i=1}^{n} I(Y_{1i}; U_{i}) \]  

or
\[ R_{0} \geq \frac{1}{n} \sum_{i=1}^{n} I(Y_{1i}; U_{i}) \].  

From (38),
\[ \frac{1}{n} H(X_{1}^{n}|S_{0}^{n}) \geq \frac{1}{n} \sum_{i=1}^{n} H(X_{i}|U_{i}) \]  

From [1, p. 578], a time-sharing uniform random variable \( Q \) distributed over \( \{1, 2, \ldots, n\} \) is introduced, which is independent of \( (X_{i}, Y_{1i}, U_{i}) \). Therefore, \( I(Y_{1i}; U_{i}) = I(Y_{1i}; U_{i}|Q = i) \) and \( H(X_{i}|U_{i}) = H(X_{i}|U_{i}, Q = i) \). These are applied to (45) and (46), which can be rewritten as
\[ R_{0} \geq \frac{1}{n} \sum_{i=1}^{n} I(Y_{1i}; U_{i}|Q = i) = I(Y_{1Q}; U_{Q}) \]  

From the chain rule, the right-hand side in (47) can be rewritten as
\[ I(Y_{1Q}; U_{Q}|Q) = I(Y_{1Q}; U_{Q}, Q) - I(Y_{1Q}; Q) \]
\[ = I(Y_{1Q}; U_{Q}) \]  

where \( I(Y_{1Q}; Q) = 0 \) in (49) because \( Y_{1Q} \) and \( Q \) are independent. The pair \( (Y_{1Q}, X_{Q}) \) have the same joint distribution as \((Y_{1}, X)\). Define \( X \equiv X_{Q}, Y_{1} \equiv Y_{1Q}, \) and \( U \equiv (U_{Q}, Q) \). Then, substitute these and (50) into (47) and (48), which can be rewritten as
\[ R_{0} \geq I(Y_{1}; U) \]  

This completes the proof of Lemma 1.

To prove the converse theorem, it is necessary to show that \( R \leq C \), if the average codeword error probability \( P^{(n)}(E) \rightarrow 0 \) for sufficiently large \( n \). First, construct a codebook \( \mathcal{U}^{n} \), where each codeword is i.i.d. and satisfies the constraint \( I(U; Y_{1}) \leq R_{0} \). The cardinality bound is the 08/30/2009
same as the one in [1, Theorem 15.8.1], which can be proved using the Fano’s inequality [1]. Let \( W \in \{1, 2, ..., 2^{nR}\} \) be the source message random variable, which carries the input message. Because \( W \) is a uniform random variable,

\[
nR = H(W)
\]

\[
\overset{(a)}{=} I(W; Y^n, S^n_0) + H(W|Y^n, S^n_0)
\]

\[
\overset{(b)}{\leq} I(W; Y^n, S^n_0) + n\varepsilon_n
\]

\[
\overset{(c)}{\leq} I(X^n; Y^n, S^n_0) + n\varepsilon_n
\]

\[
= I(X^n; Y^n|S^n_0) + I(X^n; S^n_0) + n\varepsilon_n
\]

\[
= H(Y^n|S^n_0) - H(Y^n|X^n, S^n_0) + H(X^n) - H(X^n|S^n_0) + n\varepsilon_n
\]

\[
= H(Y^n|S^n_0) - H(Y^n = X^n + Z^n|X^n, S^n_0) + H(X^n) - H(X^n|S^n_0) + n\varepsilon_n
\]

\[
\overset{(d)}{=} H(Y^n|S^n_0) - H(Z^n|X^n, S^n_0) + H(X^n) - H(X^n|S^n_0) + n\varepsilon_n
\]

\[
\overset{(e)}{=} H(Y^n|S^n_0) - H(Z^n) + H(X^n) - H(X^n|S^n_0) + n\varepsilon_n
\]

\[
\overset{(f)}{\leq} nH(Y|U) - nH(Z) + n - nH(X|U) + n\varepsilon_n
\]

\[
= n[H(Y|U) - H(Z) + 1 - H(X|U)] + n\varepsilon_n
\]

\[
\overset{(g)}{\leq} nC + n\varepsilon_n
\]

\[
R \leq C + \varepsilon_n
\]

where

- (a) follows from the definition of mutual information;
- (b) follows from the Fano’s inequality, i.e., \( H(W|Y^n, S^n_0) \leq n\varepsilon_n \) because \( p^{(n)}(E) \to 0 \);
- (c) follows from the data processing inequality;
- (d) follows from the fact that \( X^n \) is given;
- (e) follows from the fact that \( Z^n \) is independent of \( (X^n, S^n_0) \);
- (f) follows from the fact that \( H(Y^n|S^n_0) = nH(Y|U) \) because \( Y^n \) is i.i.d., and the inequality \( H(X^n|S^n_0) \geq nH(X|U) \) holds from Lemma 1; and
- (g) follows by substituting the capacity in (1) into (63).

Therefore, this completes the proof for the converse of Theorem 1.
APPENDIX C
(Proof of Theorem 2: Cut-Set Bound)

From the multiple-access cut in Fig. 4,
\[
I(X, X_1; Y, S_0) \overset{(a)}{=} H(Y, S_0) - H(Y, S_0 | X, X_1) \tag{66}
\]
\[
\overset{(b)}{=} H(S_0) + H(Y | S_0) - H(Y = X + Z, S_0 = X_1 + N_2 | X, X_1) \tag{67}
\]
\[
\overset{(c)}{=} H(S_0) + H(Y | S_0) - H(Z, N_2) \tag{68}
\]
\[
\overset{(d)}{\leq} 1 + 1 - H(Z) - H(N_2) \tag{69}
\]
\[
= 1 - H(Z) + 1 - H(N_2) \tag{70}
\]
\[
\overset{(e)}{\leq} 1 - \mathcal{H}(p) + R_0 \tag{71}
\]

where
(a) follows from the definition of mutual information;
(b) follows from the chain rule;
(c) follows from the fact that \( Y \) and \( S_0 \) are functions of \( (X, Z) \) and \( (X_1, N_2) \), respectively;
(d) follows from the fact that \( H(S_0) \leq 1 \) and \( H(Y | S_0) \leq 1 \) because the entropy of a binary random variable is upper-bounded by 1; \( H(Z, N_2) = H(Z) + H(N_2) \) because the pair \( (Z, N_2) \) are independent; and
(e) follows from (29) and \( Z \sim Ber(p) \).

From the broadcast cut in Fig. 4,
\[
I(X; Y, Y_1) \overset{(a)}{=} I(X; Y) + I(X; Y_1 | Y) \tag{72}
\]
\[
\overset{(b)}{=} I(X; Y) + H(Y_1 | Y) - H(Y_1 | Y, X) \tag{73}
\]
\[
= I(X; Y) + H(Y_1 | Y) - H(Y_1 = X + N_1 | Y, X) \tag{74}
\]
\[
\overset{(c)}{=} I(X; Y) + H(Y_1 | Y) - H(N_1 | Y, X) \tag{75}
\]
\[
\overset{(d)}{=} I(X; Y) + H(Y_1 | Y) - H(N_1) \tag{76}
\]
\[
\overset{(e)}{\leq} 1 - \mathcal{H}(p) + 1 - \mathcal{H}(\delta) \tag{77}
\]

where
(a) follows from the chain rule;
(b) follows from the definition of mutual information;
(c) follows from \( H(Y_1 = X + N_1 | X) = H(N_1) \);
(d) follows from the fact that \( N_1 \) is independent of the pair \( (Y, X) \);
(e) follows from the fact that \( I(X; Y) = 1 - \mathcal{H}(p) \), and \( H(Y_1 | Y) \leq 1 \), and \( H(N_1) = 1 - \mathcal{H}(\delta) \).

From (5), (71), and (77), the final cut-set bound for this particular channel is equal to
\[
\min \{ 1 - \mathcal{H}(p) + R_0, \ 1 - \mathcal{H}(p) + 1 - \mathcal{H}(\delta) \}. \tag{78}
\]

Therefore, this completes the proof of Theorem 2 for the cut-set bound.

08/30/2009
REFERENCES