The Maximum Achievable Throughput of a Decode-and-Forward Based Hybrid-ARQ Protocol

Ilmu Byun and Kwang Soon Kim†
Department of Electrical and Electronic Engineering, Yonsei University
134, Shinchong Dong, Seodaemun Gu, Seoul 120-749, Korea
E-mail: {dlfanv, ks.kim}@yonsei.ac.kr
† Corresponding Author

ABSTRACT

Cooperative hybrid-ARQ (HARQ) protocols have been widely studied because they are more efficient than cooperative protocols without HARQ. In [7], the throughput of a cooperative HARQ protocol based on the decode-and-forward protocol (DF-HARQ) is obtained. In this paper, the maximum achievable throughput of the DF-HARQ protocol is obtained using the asymptotic outage probability when the maximum number of transmission (M) goes to infinity. The range of the optimum worst-case coding rate \( R \) of the maximum achievable throughput is also obtained. Furthermore, we can expect that the methodology for obtaining the maximum achievable throughput in this paper also applicable to other cooperative HARQ protocols.

INTRODUCTION

To achieve more reliable communication and improve the efficiency of cooperative communications, cooperative ARQ or hybrid-ARQ (HARQ) protocols have been widely studied [1]-[3]. In [1], it was shown that the incremental relaying protocol based on the limited feedback from the destination, which can be viewed as an extension of a ARQ into a cooperative context, outperforms the fixed relaying protocol. In [2], the dynamic decode-and-forward (DDF) based ARQ protocol was proposed. It can achieve the optimum diversity-multiplexing tradeoff in a single relay channel when the maximum allowable number of transmission is greater than 1. On the other side, in [3], it was shown that an ARQ protocol can improve performance by combining it with a cooperative protocol.

There have been many further researches to develop and analyze cooperative ARQ or HARQ protocols due to their performance improvement [4]-[6]. In [4], three cooperative ARQ protocols based on the incremental relaying and the selection relaying protocols were proposed and their approximated packet error rate was obtained. In [5] and [6], cooperative HARQ protocols for multiple relays were analyzed for the Chase combining and the incremental redundancy (IR) schemes. The upper bound of the throughput of a cooperative incremental redundancy (IR) HARQ protocol was developed for a scenario where relays are close to a source [5]. However, in practical wireless systems, the relay can be selected from various locations. Thus, in [7], the performance of a cooperative IR-HARQ protocol is analyzed and the optimum initial transmission rate is also searched stepwisely according to the overall relay location. The analyzed cooperative IR-HARQ protocol in [7] is a decode-and-forward based HARQ (DF-HARQ) protocol in which an extended version of the protocol 1 in [4]. The advantages of the DF-HARQ are that it can achieve full diversity order and it is much easier to implement for practical systems.

In this paper, the maximum achievable throughput (i.e. the throughput when \( M \) goes to infinity) of the DF-HARQ protocol and the range of the optimum worst-case coding rate \( R = r/M \) is obtained for block fading channels. They are simply expressed as the terms of the ergodic capacity of each link. Furthermore, we can expect that the maximum achievable throughput of various protocols can be compared using the methodology developed in this paper.

SYSTEM MODEL AND PROTOCOL DESCRIPTION

A single relay cooperative ARQ model with three half-duplex terminals (a source, a destination, and a relay) is assumed as in [4]. The source transmits data to the destination with the aid of the relay. The destination sends one-bit feedback (ACK/NACK) to the source and the relay for retransmission. Also, the relay sends one-bit feedback to the source. We assume that ACK/NACK messages are decoded without error.

A. Protocol Description

One round of the DF-HARQ protocol is comprised of two steps. In Step 1, the source broadcasts a packet to the destination and the relay until one of them successfully decodes or the number of transmission reaches to the maximum number of transmission. If the destination
successfully decodes the packet, then the retransmission is terminated and the next round begins. Otherwise, if the destination fails to decode but the relay successfully decodes the packet, Step 2 begins. In Step 2, the relay transmits the packet to the destination until the destination successfully decodes or the number of transmission reaches the maximum number of transmission.

In the DF-HARQ protocol, the source and the relay use different codeword sets for transmission. They transmit different codeword blocks at every transmission to increase redundancy like as the incremental redundancy hybrid-ARQ protocols. Thus channel codes used in the DF-HARQ protocol is assumed to be rateless codes or rate-compatible codes.

B. Signal Model

The initial transmission rate \( r \) (b/s/Hz) of the source is fixed during one round and it is selected based on the long-term averaged signal to noise ratio (SNR). Let \( M \) be the maximum transmission number of each step. If the destination successfully decodes with \( m \) received packets, where \( m \leq M \), the effective transmission rate is \( r/m \). Especially, the effective transmission rate when \( m = M \) is defined as the worst-case coding rate \( R \), i.e. \( R = r/M \).

It is assumed that the receiver knows the channel state information perfectly. Let \( x_{\alpha} [m] \) be the \( m \)-th transmitted packet from the terminal \( \alpha \in \{s,d,r\} \), where \( s \), \( d \) and \( r \) respectively denote the source, the destination, and the relay. When a terminal transmits a packet, the \( m \)-th received packet in the terminal \( \beta \in \{s,d,r\} \setminus \alpha \) is \( y_{\beta} [m] = \sqrt{g_{\alpha,\beta} h_{\alpha,\beta}} [m] x_{\alpha} [m] + n_{\beta} [m] \),

where \( g_{\alpha,\beta} \) and \( h_{\alpha,\beta} [m] \) respectively denote the channel gain and the channel coefficient between terminal \( \alpha \) and terminal \( \beta \). In this paper, the channel is assumed to be a block fading channel with independent and identically distributed (i.i.d) complex Gaussian random variables with \( CN(0,1) \), i.e. the channels of each transmission are independent. Also, \( n_{\beta} [m] \) is the additive white Gaussian noises (AWGN) at terminal \( \beta \) with \( CN(0,N_0) \). The average received SNR of terminal \( \beta \) of \( \alpha \)-to-\( \beta \) link is

\[
\gamma_{\alpha,\beta} = g_{\alpha,\beta} E_{\alpha}/N_0
\]

Using the renewal reward theorem, the throughput of ARQ systems for the point-to-point link is obtained as [8]

\[
\eta = \frac{E[ R ]}{E[ T ]},
\]

where \( E[ R ] \) and \( E[ T ] \) are the average rate and the average number of transmission, respectively. If successful decoding is achieved until the \( M \)-th transmission, \( R = r \) b/s/Hz. If decoding is failed at the \( M \)-th transmission, \( R = 0 \) b/s/Hz. The throughput also can be used for the cooperative hybrid-ARQ protocols. In [7], \( E[ R ] \) and \( E[ T ] \) for the DF-HARQ protocol is obtained using the state transition diagram.

ASYMPTOTIC OUTAGE PROBABILITY

Let \( I_O \) be the number of orthogonal links attached to a terminal and the (vector) variables \( m = (m_1, m_2, \cdots, m_{I_O}) \) be the number of the transmission of the \( i \)-th link, for \( i = 1, \cdots, I_O \). Then the outage probability of the terminal of the IR-HARQ systems can be given by

\[
p(m,\gamma,r) = \Pr \left( \sum_{i=1}^{I_O} \sum_{t=1}^{m_i} J(\gamma_i[t]) < r \right),
\]

where \( \gamma_i[t] \) denotes the instantaneous received SNR of the \( i \)-th link at the \( t \)-th transmission. The average received SNR of the \( i \)-th link is defined as \( \gamma_i = E[\gamma_i[t]] \) and the vector variables of it denoted as \( \gamma = (\gamma_1, \gamma_2, \cdots, \gamma_{I_O}) \). For Gaussian inputs, the mutual information is given as

\[
J(\gamma_i[t]) = \log_2(1 + \gamma_i[t]).
\]

The outage probability can be simplified when \( M \) goes to infinity. First, we consider the case when \( M \) goes to infinity and \( R \) goes to zero (i.e. \( r \) has a finite value). In this case, the outage probability can be simplified as

\[
\lim_{M \to \infty, R \to 0} p(m,\gamma,r) = 0,
\]

since the sum of the mutual information goes to infinity as the number of transmission goes to infinity.

Second, we consider the case when \( M \) goes to infinity and \( R \) is greater than zero. Let \( \alpha_i = m_i/M \) be the fraction of the number of transmission incident to the maximum number of transmission. Let \( S_0 \) denote the set of the indices of \( \alpha_i \) converges to 0 and \( S_0^c \) denote the complementary set of \( S_0 \). They are given as

\[
S_0 = \{ i | \alpha_i \to 0 \text{ when } M \to \infty, \text{ for all } i = 1, \cdots, I_O \},
\]

\[
S_0^c = \{ i | \alpha_i \to 0 \text{ when } M \to \infty, \text{ for all } i = 1, \cdots, I_O \}.
\]

Then the outage probability can be interpreted as

\[
p(m,\gamma,r) = \Pr \left( \sum_{i \in S_0} \alpha_i J(\gamma_i) + \sum_{i \in S_0^c} \alpha_i J(\gamma_i) < R \right),
\]

where \( J(\gamma_i[t]) = \frac{1}{m_i} \sum_{t=1}^{m_i} J(\gamma_i[t]) \) is the sample mean of the mutual information. The mutual information \( J(\gamma_i[t]) \) is an i.i.d random variable with a finite mean and a
variance. Thus the strong law of large numbers can be used for obtaining the asymptotic outage probability when \( M \) goes to infinity. Furthermore, the summation on \( S_0 \), \( \sum_{i \in S_0} \alpha_i C(\gamma_i) \), converges to 0 when \( M \) goes to infinity because all \( \alpha_i \in S_0 \) goes to 0. Thus the asymptotic outage probability converges to

\[
\lim_{M \to \infty, R > 0} p(m, \gamma, r) = \Pr \left( \sum_{i \in S_0} \alpha_i C(\gamma_i) < R \right),
\]

where \( C(\gamma_i) = E \left[ J(\gamma_i[t]) \right] \) is the real mean of \( J(\gamma_i[t]) \). For the Gaussian inputs, \( C(\gamma_i) \) is obtained as

\[
C(\gamma_i) = \int \log_2(1 + \gamma_i[t]) \frac{1}{2 \sigma^2} \exp\left(-\frac{\gamma_i[t]}{2 \sigma^2}\right) d\gamma_i[t] = \frac{\exp(1/\gamma_i)}{\ln(2)} \text{Ei}\left(-\frac{1}{\gamma_i}\right),
\]

where \( \sigma^2 = 0.5 \) and \( \text{Ei}(\cdot) \) denotes the exponential integral [10]. From (5) and (6), we can see that there is no random variable. Thus it can be given by the unit step function as

\[
\lim_{M \to \infty, R > 0} p(m, \gamma, r) = U \left( R - \sum_{i \in S_0} \alpha_i C(\gamma_i) \right).
\]

Here the unit step function \( U(\cdot) \) is defined slightly differently as

\[
U(t - t_0) = \begin{cases} 1, & \text{if } t > t_0, \\ 0, & \text{if } t \leq t_0. \end{cases}
\]

The asymptotic outage probabilities of (5) and (7) can be straightforwardly applied to the outage probabilities of the destination and the relay of the DF-HARQ protocol. When \( M \) goes to infinity and \( R \) goes to zero, the outage probabilities also goes to zero. When \( M \) goes to infinity and \( R \) is greater than zero, they are defined as follows. The outage probabilities of the destination and the relay in Step 1 can be respectively given by \( p(m_1, \gamma_{sd}, r) \) and \( p(m_2, \gamma_{sr}, r) \) since they have one link. In the DF-HARQ protocol, \( m_1 \) is equal to \( m_2 \) because the source transmits data to the relay and the destination in a broadcast manner. For notational simplicity, they are denoted as \( p_1(m_1) \) and \( p_2(m_2) \) when they cause no ambiguity. The asymptotic outage probabilities of them are given by

\[
\lim_{M \to \infty, R > 0} p_a(m_a) = U \left( \alpha_a^* - \alpha_a \right),
\]

where \( \alpha_a^* = R/C(\gamma_{sa,d}) \) and \( \alpha_a^2 = R/C(\gamma_{sa,r}) \), for \( a = 1, 2 \). The outage probability of the destination in Step 1 can be given by \( p(m, \gamma, r) \), where \( m = (m_2, m_3) \) and \( \gamma = (\gamma_{sd}, \gamma_{sr}) \), since the links from the source and the relay are orthogonal and Step 2 begins only when the relay successfully decodes. It is also simplified as \( p_3(m_2^*, m_3) \) for notation simplicity. The asymptotic outage probability of \( p_3 \) is given by

\[
\lim_{M \to \infty, R > 0} p_3(m_2^*, m_3) = U \left( \alpha_3^* - \alpha_3 \right),
\]

where \( \alpha_3^* = (R - \alpha_2^2 C(\gamma_{sr,d}) - C(\gamma_{sr,r})). \)

**THE MAXIMUM THROUGHPUT OF THE DF-HARQ PROTOCOL**

Previously, in [7], the average number of transmission of the DF-HARQ protocol was obtained as

\[
E[T] = 1 + \sum_{m_1 = 1}^{M-1} p_1(m_1) p_2(m_2) + \sum_{m_1 = 1}^{M} p_1(m_1) q_2(m_2) \left( 1 + \sum_{n=1}^{M-1} p_3(m_1, n) \right).
\]

The average rate of transmission was also obtained as

\[
E[R] = r(1 - p_{out}),
\]

where

\[
p_{out} = p_1(M) p_2(M) + \sum_{m_1 = 1}^{M} p_1(m_1) q_2(m_2) p_3(m_1, n).
\]

In the above equations, \( q_1(m_1) \) and \( q_2(m_2) \) denote the probabilities of successful decoding corresponding to \( p_1(m_1) \) and \( p_2(m_2) \), respectively. They can be rewritten as \( q_a(m) = p_a(m - 1) - p_a(m) \), for \( a = 1, 2 \), and \( q_3(m, n) = p_3(m, n - 1) - p_3(m, n) \). By substituting (10) and (11) into (1), the throughput of the DF-HARQ protocol can be obtained.

**Theorem 1.** The maximum achievable throughput (i.e. the throughput when \( M \to \infty \)) of the DF-HARQ protocol is obtained as

\[
C_{DF-HARQ} = \begin{cases} C(\gamma_{sa,d}), & \text{if } C(\gamma_{sd,d}) \geq C(\gamma_{sr,r}) \text{ and } 0 < R \leq C(\gamma_{sd,d}), \\ \frac{C(\gamma_{sa,d}) + C(\gamma_{sa,r}) - C(\gamma_{sr,r})}{C(\gamma_{sa,d}) + C(\gamma_{sa,r}) - C(\gamma_{sr,r})}, & \text{if } C(\gamma_{sd,d}) < C(\gamma_{sr,r}) \text{ and } 0 < R \leq U, \\ 0, & \text{otherwise,} \end{cases}
\]

where \( C(\gamma_{a,b}) \) denotes the ergodic capacity of the \( \alpha \)-to-\( \beta \) link and \( U = \min \left( C(\gamma_{sr,r}), 1 - C(\gamma_{sd,d})/C(\gamma_{sr,r}) \right). \)

**Proof:** See Appendix.

From Theorem 1, we can see that the maximum achievable throughput is constant under the given constraints.
This is because the number of transmissions linearly increases as \( r \) increases when \( M \) goes to infinity and \( R \) is greater than 0.

Under the one-dimensional linear relay network model, the fixed DF protocol has its best performance when the relay is close to the source and its performance is asymmetric according to the relay location. Whereas, from Corollary 1, we can see that the maximum achievable throughput of the DF-HARQ protocol where the relay is in the middle greater than that of other regions and it is symmetric according to the relay location. However, the region of the OWCR of the DF-HARQ is asymmetric.

Figures 1 and 2 respectively show the maximum throughput and the OWCR according to \( M \). The OWCR is searched stepwisely over \( 0 \sim 12 \) (b/s/Hz) with 0.05 step size. The maximum achievable throughput is corresponding to the OWCR. The simulation environment is as follows. The simulation is performed under the one-dimensional linear relay network model. Let the distance between the source and the destination be 1 and the distance between the source and the relay be \( l \). We assume that channel is assumed to be a block fading, and the channel gain of each link to be \( g_{s,d} = 1 \) for the source-to-destination link, \( g_{s,r} = l^{-a} \) for the source-to-relay link, and \( g_{r,d} = (1 - l)^{-a} \) for the relay-to-destination link, where \( a \) is set to 4. The source and the relay transmit their packets with the same power. From Figure 1, it is seen that the maximum throughput of a finite \( M \) approaches to that of infinite \( M \) as \( M \) increases. When \( M \) is small, the maximum throughput increases very sharply as \( M \) increases. As \( M \) increases further, the maximum throughput becomes almost a constant. Figure 2 shows that \( R \) for various \( M \) is smaller than the upper bound of \( R^* \) when \( M \) goes to infinity. Thus, the search space for obtaining the OWCR can be reduced using the analysis result.

**CONCLUSION**

In this paper, the maximum achievable throughput of the DF-HARQ protocol is analytically obtained using the state transition diagram in [7]. Since the maximum achievable throughput is expressed as the ergodic capacity of each point-to-point link, we can easily obtain it for various input and channel. Also, it is seen that the best maximum achievable throughput is obtained where the relay is in the middle and the maximum achievable throughput is symmetric under the one-dimensional linear relay network model.

**APPENDIX**

In the first case when \( M \) goes to infinity and \( R \) goes to 0, \( p_{out} \) goes to 0 since the outage probability of each terminal goes to 0 as shown in the previous section. Thus the throughput converges as

\[
\lim_{M \to \infty, R \to 0} \frac{r}{E[T]} = \frac{r}{E[T]}. \tag{13}
\]

Intuitively, we can see that the increase of the initial transmission rate \( r \) induces the increase of \( E[T] \). For the block fading channel, the increase of \( E[T] \) implies the increase of the diversity gain. Thus it is easily seen that, in the first case, the throughput \( \eta \) increases as \( r \) increases. In the second case when \( M \) goes to infinity and \( R \) greater than 0, \( p_{out} \) also goes to 0 if \( R \) is smaller than a certain
value. The initial transmission rate \( r = R/M \) in the second case is greater than that of the first case. Thus we can see that

\[
\lim_{M \to \infty, R > 0} \eta \geq \lim_{M \to \infty, R = 0} \eta,
\]

when \( R \) is smaller than a certain value. The maximum achievable throughput in the second case is as follows.

Let \( m_1^* \) and \( m_2^* \) respectively be the number of transmission for the successful decoding of the destination and the relay in Step 1. Here \( m_a^* \) means the minimum number of transmission for \( p_a(m) = 0 \), for \( a = 1, 2 \). Let \( m_3^* \) be the number of transmissions of the successful decoding at the destination in Step 2. Then the asymptotic outage probabilities, (8)-(9), can be interpreted as a discrete delta function which is 1 when \( a = 2 \) and 0 otherwise.

Thus, the average rate becomes 0. Thus, (15) can be rewritten as

\[
\lim_{M \to \infty, R > 0} \eta = \begin{cases}
  \frac{R}{m_1^*}, & \text{if } \alpha_1^* \geq \alpha_2^*, \alpha_1^* \leq 1, \\
  \frac{m_2^* + m_3^*}{r}, & \text{if } \alpha_1^* < \alpha_2^*, \alpha_2^* \leq 1, \alpha_3^* \leq 1, \\
  0, & \text{otherwise},
\end{cases}
\]

(17)

Finally, the throughput can be obtained by substituting (14) and (16) into (1), the throughput of the DF-HARQ protocol can be obtained as

\[
\lim_{M \to \infty, R > 0} \eta = \begin{cases}
  \frac{R}{m_1^*}, & \text{if } \alpha_1^* \geq \alpha_2^*, \alpha_1^* \leq 1, \\
  \frac{m_2^* + m_3^*}{r}, & \text{if } \alpha_1^* < \alpha_2^*, \alpha_2^* \leq 1, \alpha_3^* \leq 1, \\
  0, & \text{otherwise}.
\end{cases}
\]

(17)

The maximum achievable throughput can be obtained by substituting the values of \( \alpha_1^* \) and \( \alpha_2^* \) into (17).

ACKNOWLEDGEMENT

This research was supported by the Ministry of Knowledge Economy, Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA- 2009-C1090-0902-0010)

REFERENCES


