A Queueing Model of an Adaptive Type-I Hybrid-ARQ with a Generalized Markovian Source and Channel

Jun-Bae Seo, Seung-Que Lee and Nam-Hoon Park
Department of Wireless System Research, ETRI, Deajeon Korea
e-mail: jbseo, sqlee, nhpark@etri.re.kr

Hyong-Woo Lee
Department of Electronics and Information Engineering, Korea University
e-mail: hwlee@korea.ac.kr

Choong-Ho Cho
Department of Computer and Information Science, Korea University
e-mail: chcho@korea.ac.kr

Abstract

In this paper, the queueing performance of an adaptive type-I hybrid automatic repeat request(ARQ) with a generalized Markovian source is studied for a TDMA/FDD system over a Markovian channel. In the adaptive type-I hybrid ARQ of our interest, the payload size for information in a packet(or ARQ block) and the coding rate of ARQ block vary, according to the channel state, e.g., high coding rate and small payload size for information in a bad state of channel. We assume that an arriving message over a medium access control(MAC) layer consists of a number of information units of fixed size. We analyze the delay of a message, which is composed of its queueing delay and its transmission time, and the message dropping probability. In the numerical examples, we have shown that the correlation of either traffic source or radio channel affects the queueing performance in a similar way, due to its Markovian structure. The analytical work presented here can be used to control physical parameters, such as coding rate, in order to meet QoS requirements in a higher layer which is over the MAC layer.

Keywords—Type-I Hybrid ARQ.

1 Introduction

The provision for quality-of-service(QoS) of data service in a wireless packet mobile network makes it necessary to consider QoS requested from a wireline network connected to an access point of a wireless mobile network. In order to incorporate QoS of a wireline domain into that of a wireless domain seamlessly, it is required to appropriately control physical parameters assigned to each user, such as channel coding rate and modulation schemes to transmit over unreliable and time-varying radio channels. As a basic configuration of this scheme, the code rate of an error correcting code is adaptively chosen in order to compensate for the fluctuation in channel error statistics for maximizing system throughput. The previous studies[1]-[3] have been focused on the system throughput of this scheme; However, its queueing performance has been seldom studied, although it is strongly related to QoS parameters, such as delay and loss requirements. In [4], the queueing response for a TDMA/FDD system with fixed coding rate is investigated with measurement-based fading statistics of log-normal and multipath fading. In [5][6], it is studied to find effective capacity of a type-I hybrid ARQ scheme for supporting delay constraint and loss rate by fluid-flow approximation. Though their works provide an effective capacity given QoS as a closed form solution, in general, fluid flow analysis gives a lower bound for the quantity of interest by ignoring the dynamics of a point process. In our model, messages of variable size arrive over a MAC layer and are served within a MAC layer by ARQ block of fixed-size, of which the information content excluding redundant bits for error correction varies according to the channel state. This assumption can reflect a fragmentation operation. The message delay is obtained by considering its queueing delay and transmission delay which is the time to transmit the entire information contained in a message. The organization of this paper is as follows. In section 2, the system is described in detail. In section 3, the queue is analyzed and the message delay is derived. The numerical examples is given in section 4. The concluding remarks are given in section 5.

2 System Description

2.1 Source Model

A traffic source is in one of $M$ states and the sequence of source states is assumed to be a discrete-time Markov chain whose one step state transition probability matrix, $P$, is as follows.

$$ P = \begin{bmatrix}
        p_{0,0} & p_{0,1} & \cdots & p_{0,M-1} \\
        p_{1,0} & p_{1,1} & \cdots & p_{1,M-1} \\
        \vdots & \vdots & \ddots & \vdots \\
        p_{M-1,0} & \cdots & \cdots & p_{M-1,M-1}
    \end{bmatrix} \quad (1) $$

where $p_{ij}$ is the transition probability from state $i$ to state $j$ in one frame. Let $\pi_{s,j}$ be the steady state probability that the traffic source is in $i$-th state during a frame. Further, $\pi_{s}$ is defined as

$$ \pi_{s} = [\pi_{s,0}, \pi_{s,1}, \ldots, \pi_{s,M-1}] \quad (2) $$

satisfies the following equations.

$$ \pi_{s} = \pi_{s} P \quad \text{and} \quad \pi_{s} e = 1 \quad (3) $$

Henceforth, we assume that the state transition occurs just after each frame boundary. The number of arriving messages produced in each state during a frame over MAC layer is an identically and independently distributed(i.i.d) random variable with mean $\nu \mu^s(\nu)$, where $\nu$ denotes $\nu$-th source state. A message consists of a number of information units, e.g., bits or bytes, which is also characterized by i.i.d random variable with mean $\nu \beta^s(\nu)$. Let $G^{(\nu)}_m(z)$ and $G^{(\nu)}_n(z)$ respectively denote the probability generating function(PGF) of the number of arriving message during a frame and that of information units in a message in $\nu$-th state. Then, the message arrival process in a frame at $\nu$-th state forms
a compound process. Its PGF denoted by $G_B^{(\nu)}(z)$ can be expressed as

$$G_B^{(\nu)}(z) = G_m^{(\nu)} \left( G_b^{(\nu)}(z) \right)$$  \hspace{1cm} (4)$$

Especially, we assume that no message is produced in the ‘0’-th state, which is ‘OFF’-state, i.e., $G(0) = 1$. In other states, at least one message is produced.

### 2.2 Channel Model and Service Process

The radio channel is in one of $N$ states and the sequence of channel states is assumed to be a discrete-time Markov chain whose transition probability matrix, $R$, is given by

$$R = \begin{bmatrix} r_{0,0} & r_{0,1} & \cdots & r_{0,N-1} \\ r_{1,0} & r_{1,1} & \cdots & r_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N-1,0} & \cdots & \cdots & r_{N-1,N-1} \end{bmatrix}$$  \hspace{1cm} (5)$$

where $r_{ij}$ denotes the transition probability from state $i$ to $j$ in one slot and $r_{ij} = 0$, for $|i - j| > 1$. Let $\pi_{c,i}$ be the steady state probability that channel state for the tagged user whose buffer we are analyzing is $i$. Then, $\pi_c$ defined as

$$\pi_c = [\pi_{c,0}, \pi_{c,1}, \ldots, \pi_{c,N-1}]$$  \hspace{1cm} (6)$$

satisfies the following equations

$$\pi_c = \pi_c R^s \quad \text{and} \quad \pi_c e = 1$$  \hspace{1cm} (7)$$

where $[\cdot]^s$ means the $s$-th power of $[\cdot]$ and $r_{ij}^s$ denotes the $i$-th row and $j$-th column element of $R^s$.

When the channel is in state $i$, the bit error probability is $\epsilon_i$ ($\epsilon_0 > \epsilon_1 > \ldots > \epsilon_{N-1}$). The number of bits to be encoded by a proper block code, the number of block-coded bits sent to a modulator and the number of correctable bits by a block code used in the $l$-th state will be denoted by $k_l, l_1$ and $l_2$, respectively. Consequently, $n - k_l$ is the number of redundant bits for error correction in the $l$-th channel state. It may be assumed that $l_2$ consists of the overhead of MAC layer, $C_{\alpha}$ bits, the payload for the information, $C_l$, and a cyclic redundancy check(CRC) code of $C_{\beta}$ bits, i.e., $k_l = C_{\alpha} + C_l + C_{\beta}$. When the amount of the remaining information units in the queue is less than $C_l$, some dummy bits are appended to fulfill a payload. Otherwise, a message is fragmented into $C_l$ and it is transmitted as an ARQ block. We assume that a strong CRC is applied so that all error patterns are detected. The probability that a packet received will contain an uncorrectable error, $P_{e,l}$ in $l$-th channel state by a block code, ($n, k_l, l_1$), is given by

$$P_{e,l} = \frac{n}{\sum_{j=l_2+1}^{n} \binom{n}{j} \epsilon_l^j (1 - \epsilon_l)^{n-j}} \quad \text{and} \quad P_l = 1 - P_{e,l}$$  \hspace{1cm} (8)$$

where $P_l$ is the probability that a packet transmission is successful in the $l$-th channel state.

### 2.3 TDMA/TDD System

In the TDMA/TDD system of our interest, time is divided into contiguous slots. The slot duration is $\tau$ seconds and equals to the packet(ARQ block) transmission time. One frame consists of downlink and uplink, whose lengths are respectively $d \cdot \tau$ and $u \cdot \tau$. Here, $d$ and $u$ are, respectively, the number of slots in a down- and uplink. For a user, one slot in downlink of a frame is assigned such that the information of a user is periodically transmitted on it, e.g., every frame. A packet is transported by Selective Repeat ARQ scheme in which ACK/NACK messages for a packet transmitted in the downlink of the $n$-th frame always arrive without errors in the uplink of the $n$-th frame. Further, we assume that the transmitter always perfectly knows the channel state of the downlink in the next frame by a channel estimation and/or prediction scheme of a TDMA/TDD system[7], or there is a dedicated control channel in the uplink, which is assigned to a user for reporting the quality of radio channel state.

### 3 Analysis

#### 3.1 Queueing Analysis

As previously mentioned, the frame length is equal to $s(= d + u)$ slots. The packet transmission occurs at the beginning of a slot in downlink, provided the arriving messages just after the transmission time are queued, which is often called the gated service. An ACK/NACK message will be received before the transmission time of the next frame. One can observe that each transmission period is $s$ slots and is equal to the frame length. We assume that the channel state changes just prior to a slot boundary. Accordingly, between two consecutive transmission times, the channel state changes by $s$-step transition. According to each channel state, the transmitter chooses a coding rate and the payload size of a packet varies in accordance with the channel state, due to a fixed modulation scheme. From the transmitter buffer, information is released only when an ACK message is received for its last transmission attempt. In a MAC layer header within a packet, the amount of information carried in a packet is indicated so that the receiver can correctly organize a message. The state space of the embedded Markov chain for the system can be described by $\{(q_j^{s,i}, \alpha = k, j^o = j, i^o = i) : 0 \leq k \leq \infty, j^o \in \{0, 1, \ldots, M - 1\} \text{ and } i^o \in \{0, 1, \ldots, N - 1\}\}$ where $k, j^o$ and $i^o$ denote the number of information units in the queue, the states of source and channel, respectively. The number of information units in the queue at the end of a transmission slot of the $n$-th frame, i.e., an embedded point, with $j$-th source and $i$-th channel state will be denoted by $q_j^{n,i}$. Further the steady state probability that $k$ information units are in the queue, $q_j^{i,k}$, is defined as

$$q_j^{i,k} = \lim_{n \to \infty} \Pr[q_j^{n,i} = k]$$  \hspace{1cm} (9)$$

Also, its PGF is defined as

$$Q_j^{i}(z) = \sum_{k=0}^{\infty} q_j^{i,k} z^k$$  \hspace{1cm} (10)$$
or in an $MN$ dimensional column vector form,

$$Q(z) = \begin{bmatrix} Q_{0,0}(z) & Q_{0,1}(z) & \cdots & Q_{0,N-1}(z) \\
Q_{1,0}(z) & Q_{1,1}(z) & \cdots & Q_{1,N-1}(z) \\
\vdots & \vdots & \ddots & \vdots \\
Q_{M-1,0}(z) & Q_{M-1,1}(z) & \cdots & Q_{M-1,N-1}(z) \end{bmatrix}^T$$

(11)

where the superscript $T$ denotes matrix transpose. The evolution of the queue length can be obtained as the following state balance equations.

If $k = 0$,

$$q_{j,i}[0] = \sum_{\nu} p_{\nu,i} q^{(\nu)}_{\nu,0} \sum_{l=0}^{N-1} t_{l,i}^{(\nu)} \left( q_{\nu,l}[0] + C_l p_{\nu,l}[i] \right)$$

(12)

where $a^{(\nu)}_k$ is the probability that $k$ information units arrive in $s$ slot when traffic source is in the $\nu$-th state. Further, $a^{(0)}_0 = 1$ and $a^{(0)}_k = 0$, for $k \geq 1$, because no message is produced in the '0'(OFF) state and $a^{(\nu)}_0$ is equal to zero for $\nu \geq 1$.

And if $k \geq 1$,

$$q_{j,i}[k] = \sum_{\nu} p_{\nu,i} G^{(\nu)}(z) \sum_{l=0}^{N-1} t_{l,i}^{(\nu)} \left\{ P_l a^{(\nu)}_l \sum_{\eta=0}^{C_l} q_{\nu,\eta,l}[k + C_l - \eta] + a^{(\nu)}_l q_{\nu,l}[0] \right\}$$

(13)

Substituting (12) and (13) into (10) and after some lengthy manipulation, one can obtain

$$Q_{j,i}(z) = \sum_{\nu} p_{\nu,i} G^{(\nu)}(z) \sum_{l=0}^{N-1} t_{l,i}^{(\nu)} \left[ (1 - P_l) + P_l z^{-C_l} \right] Q_{\nu,l}(z)$$

$$= \sum_{\nu} p_{\nu,i} G^{(\nu)}(z) \sum_{l=0}^{N-1} t_{l,i}^{(\nu)} P_l (1 - z^{-C_l}) \sum_{\eta=0}^{C_l} q_{\nu,l}[\eta]$$

(14)

for $j \in \{0, 1, ..., M-1\}$ and $i \in \{0, 1, ..., N-1\}$.

As a matrix form of (14), one can readily write

$$\begin{bmatrix} I - [P^T A(z)] & [R^T E(z)]^T \end{bmatrix} Q(z) = \begin{bmatrix} P^T A(z) & [R^T I - E(z)] \end{bmatrix} \tilde{Q}_0$$

(15)

where $\otimes$ denotes the Kronecker product of two matrices and $I$ is an identity matrix of $(MN) \times (MN)$ dimension. $A(z)$ and $E(z)$ are an $(M \times M)$ diagonal matrix whose diagonal elements are $G^{(\nu)}_z(\nu)$ and an $(N \times N)$ diagonal matrix with the $l$-th element on the diagonal equal to $e_l(z) = [(1 - P_l) + P_l z^{-C_l}]$, respectively. In (15), the column vector, $\tilde{Q}_0$, is given by

$$\tilde{Q}_0 = [\tilde{q}_{0,0} \ \tilde{q}_{0,1} \ \cdots \ \tilde{q}_{0,0,N-1} \ \tilde{q}_{1,0} \ \cdots \ \tilde{q}_{1,0,N-1} \ \cdots \ \tilde{q}_{M-1,0} \ \cdots \ \tilde{q}_{M-1,0,N-1}]^T$$

(16)

with

$$\tilde{q}_{j,i} = \sum_{\nu} q_{j,i}[\nu] \ \forall i, j, \ \tilde{q}_{0,j} \neq 0, \forall j \text{ and } \tilde{q}_{k,j} = 0, i \geq 1, \forall j$$

(17)

In order to determine the unknown boundary terms, $\tilde{q}_{0,i}$, $\forall i$, all poles of $Q(z)$ should be in a closed unit disk for a stable system providing that $Q(z)$ is analytic in a closed unit disk. We follow the similar approach as in [8][9]. The following diagonalization of the matrix $P^T A(z) \otimes R^T E(z)$ may facilitate finding these poles.

$$P^T A(z) \otimes R^T E(z) = G(z) \Lambda(z) G^{-1}(z)$$

(18)

where $\Lambda(z)$, $G(z)$ and $G^{-1}(z)$ are $(MN) \times (MN)$ dimensional matrices, each of which is given by

$$\Lambda(z) = \text{diag}[\lambda_0(z), \lambda_1(z), ..., \lambda_{MN-1}(z)]$$

$$G(z) = [g_{0,0}(z), g_{1,0}(z), ..., g_{MN-1,0}(z)]$$

$$G^{-1}(z) = [g_{0,0}(z), g_{1,0}(z), ..., g_{MN-1,0}(z)]^T$$

(19)

where $g_{0,0} \text{ and } h_{0,0}$ are respectively the left column and right row eigenvector corresponding to $\lambda_l(z)$ for $\ell \in \{0, 1, ..., MN-1\}$.

Define $\xi(z)$ and $\zeta(z)$ as

$$\xi(z) = [g_{0,0}(z), g_{1,0}(z), ..., g_{MN-1,0}(z)]$$

$$\zeta(z) = [g_{0,0}(z), g_{1,0}(z), ..., g_{MN-1,0}(z)]^T$$

(20)

By spectral decomposition, one can rewrite (18) as

$$P^T A(z) \otimes R^T E(z) = \sum_{\ell=0}^{MN-1} \lambda_{\ell}(z) g_{\ell}(z) h_{\ell}(z)$$

(21)

Then, (21) can be simplified as

$$Q(z) = \sum_{\nu=0}^{\nu+1} \left[ P^T A(z) \otimes R^T E(z) \right]^{\nu+1} E(z) \tilde{Q}_0$$

$$= \sum_{\nu=0}^{MN-1} \sum_{\ell=0}^{MN-1} \lambda_{\nu,\ell+1}(z) g_{\ell}(z) h_{\ell}(z) E(z) \tilde{Q}_0$$

(22)

with

$$\lambda_{\nu,\ell}(z) = \lambda_{\ell}(z)/(1 - \lambda_{\ell}(z))$$

and $E(z) = I \otimes E(z)^{-1} - I$

Let the characteristic function of the system, $\Delta(z)$, be defined as

$$\Delta(z) = \prod_{\ell=0}^{MN-1} \left[ 1 - \lambda_{\ell}(z) \right]$$

(23)
The poles of (22) are the roots of this characteristic function, which is separately solved for each term. Since the system was described by an irreducible Markov chain, one must have \( \lambda_0(1) = 1 \). Let \( z_{\ell}^j \), for \( |z_{\ell}^j| < 1 \) and \( 1 \leq \ell \leq MN - 1 \), denotes the zero of (23). Since \( Q(z) \) is analytic at these zeros, one can obtain the following boundary equation.

\[
h_{\ell}(z_{\ell}^j)E(z_{\ell}^j)Q_0 = 0 \tag{24}
\]

With the following relation, one can find another boundary equation for the root \( z_0^1 = 1 \).

\[
\lim_{z \to 1} e \cdot Q(z) = 1 \tag{25}
\]

Using L'Hopital’s rule for (22), one can obtain

\[
- \lambda'_0(1) = e \cdot q_0(1)h_0(1)E'(1)Q_0 = e \cdot E'(1)Q_0 \tag{26}
\]

The equation (26) is rewritten by

\[
\sum_{\nu=0}^{N-1} C_\nu P_{\nu} \pi_{\nu,0} \pi_{\nu,0} - \sum_{\nu=0}^{M-1} \sum_{\nu=0}^{N-1} \frac{N-1}{\nu} C_\nu P_{\nu} q_{0,\nu}[0] \tag{27}
\]

Whenever the left-hand side of (27) is greater than zero, which implies a stable system, one can determine all the unknown boundary terms, \( \tilde{q}_{0,j}, \forall j \), with (24) and (27). The system utilization, \( \rho \), can be obtained from (27),

\[
\rho = \frac{\sum_{\nu=0}^{M-1} \sum_{\nu=0}^{N-1} \frac{N-1}{\nu} C_\nu P_{\nu} \pi_{\nu,0}}{\sum_{\nu=0}^{M-1} \sum_{\nu=0}^{N-1} C_\nu P_{\nu} \pi_{\nu,0}} \tag{28}
\]

### 3.2 Message Delay Analysis

For analyzing the mean message delay, we observe a tagged message arriving in an arbitrary frame. The delay of the tagged message involves its queuing delay and the transmission delay. We assume that it takes one slot time to transmit the remaining information of the tagged message, when the size of the remaining information is less than or equal to the payload of a packet. Let \( v_{1,j} \) denote the sum of the number of information units in the queue, the number of information units in the messages which have arrived just before the arrival of the tagged message at the \( n \)-th frame, and the number of information units contained in the tagged message, given that the source state is not the 0(OFF)-th state with the \( j \)-th channel state. Henceforth, we assume a two-state ON-OFF source for simplicity of analysis. The mean delay for transmitting the tagged message, when the channel state in an arbitrary frame is \( j \), denoted by \( D_j \), can be expressed as

\[
D_j = \sum_{k=1}^{\infty} d_j[k] \Pr[v_{1,j} = k] \tag{29}
\]

where \( d_j[k] \) is the expected number of frames needed to transmit the tagged message, conditioned on that the channel state of the frame is in \( j \). If \( 1 \leq k \leq C_t \), \( d_j[k] \) will be 1 with the probability \( P_j \). When transmission fails with the probability \( 1 - P_j \), it becomes 1 plus \( d_j[k] \) for retransmission, which is the expected number of frames needed to transmit \( k \) information units in the queue at the next channel state \( l \) with the probability \( r_{j,l} \).

\[
d_j[k] = 1 \cdot P_j + \left\{ 1 + \sum_{l=0}^{N-1} r_{j,l} d_l[k] \right\} (1 - P_j) \tag{30}
\]

And if \( k > C_t \), \( d_j[k] \) depends on (30) as well as the remaining information units in the queue after successful transmission.

\[
d_j[k] = 1 + \sum_{l=0}^{N-1} r_{j,l} \left\{ (1 - P_j) d_l[k] + P_j d_l[k - C_t] \right\} \tag{31}
\]

Then, the mean delay of a tagged message, \( D \), is obtained by

\[
D = s \cdot \tau \left[ \sum_{j=0}^{N-1} D_j \pi_{c,j} - 1 \right] + \tau \tag{32}
\]

In the above, it is considered that the last frame to transmit the remaining information is subtracted and just one slot time is added. For obtaining \( \Pr[v_{1,j} = k] \) in (29), define PGF of \( v_{1,j} \) as

\[
V_{1,j}(z) = \frac{1}{\pi_{s,1}} W(z) \sum_{i=0}^{N-1} Q_{1,i}(z) r_{i,j}^\tau \tag{33}
\]

where \( W(z) \) is the PGF of the random variable, \( w \), representing the sum of the number of information units in the messages arrived during the frame just before the arrival of the tagged message and the number of information units in the tagged message. Referring to p.21 in [11], \( W(z) \) can be obtained by

\[
W(z) = \frac{1 - G_B(z)}{\mathcal{M}(1 - G_B(z))} G_b(z) \tag{34}
\]

Taking the inverse \( \mathcal{Z} \)-transform of (33) and with (29)-(35), the mean message delay of (33) can be obtained. We assume that a message is dropped when its size is greater than the remaining queue size. Providing that the permissible message loss in a finite system is very small, \( P_d \) can be obtained as an approximate upper bound by capturing the tail distribution of an infinite system.

\[
P_d = \sum_{k=0}^{L} \sum_{i=0}^{1} q_{0,i}[k] \sum_{j=1}^{L-k-j} a_{i,L-k-j}^* + \sum_{k=L+1}^{\infty} \sum_{i=0}^{1} q_{0,i}[k] \tag{35}
\]

where \( a_{i,j}^* \) is the probability that a message with \( j \) information units arrives.

### 4 Numerical Examples

In the numerical examples, two-state Markovian source and channel channel are used, i.e., \( M = N = 2 \). The slot duration is 1 \( m\text{sec} \) and a frame consists of 6 slots. We assume that the number
of arriving messages of ‘ON’ state in a frame is assumed to be geometrically distributed with mean, $\overline{m}$ and $\overline{b}$ as follows:

$$\Pr[\text{n message arrivals in a frame}] = \frac{1}{\overline{m}} \left( 1 - \frac{1}{\overline{m}} \right)^{n-1}, \quad 1 \leq \overline{m}, 1 \leq n$$

(36)

The distribution of the number of information units in a message is equal to (36), if $\overline{m}$ is replaced by $\overline{b}$. Their PGFs are respectively given by $G_m(z) = z/(\overline{m} - (\overline{m} - 1)z)$ and $G_b(z) = z/(\overline{b} - (\overline{b} - 1)z)$. Thus, each term of (33) can be expressed as a rational function of $z$, which is numerically inverse $Z$-transformable.

5 Conclusion

In this paper, we analyzed a queueing system of an adaptive type-I hybrid ARQ scheme with a generalized Markovian source and channel model in a TDMA/TDD system. We assume that a message of variable size is transmitted by a fixed-size ARQ block whose payload size varies in accordance with channel state. The mean message delay is obtained, by considering its queueing delay and its own transmission delay. The message dropping probability is also derived from an infinite queue size assumption as an approximate upper bound. Though we can not provide a closed form solution to yield an appropriate coding rate for given QoS requirements, we show an analytic method to control physical parameters to meet QoS requirements of a higher layer over a MAC layer, by observing queueing performance. Our analysis may be also a preliminary work for studying a queueing performance of an adaptive coding and modulation scheme in a TDMA system.

Acknowledgments

This work was partially supported by Korea Research Foundation Grant(KRF-2004-041-D00457).

References