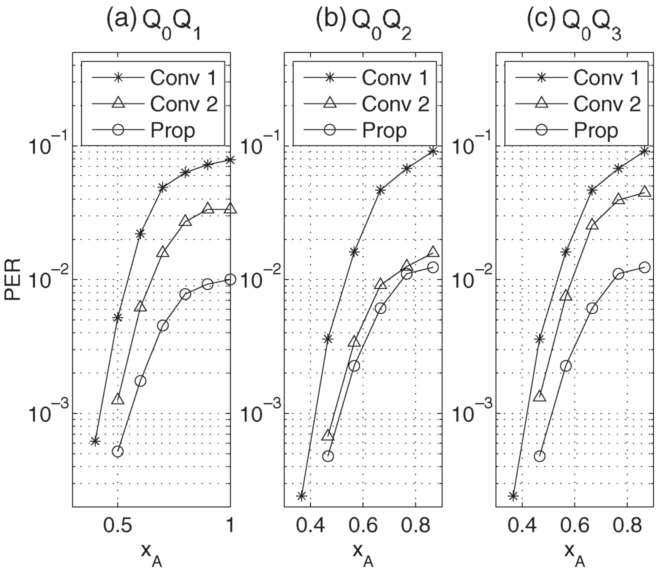

Fig. 7. Coded PER along different paths when the average SNR per BS at the three-cell boundary is set to 3.5 dB in the PA channel model with $N_T = 1$.

VI. Conclusion

We have proposed an MC method for OFDM-based broadcast services, where cells are divided into three CGs, an OFDM-symbol block is partitioned into multiple subblocks, and a combination of a CG and an antenna is assigned to each subblock for the transmission of two TD code branches. For the proposed method, the uncoded BER and the upper bound on the PEP are derived to predict the performance without lengthy simulations. With the analytically derived results, we can choose the proper CG and antenna-assignment rule for the proposed method. Simulation results also reveal that the proposed method can uniformly improve the cell-boundary performance and double the cell coverage at the same power budget without a significant increase in complexity. Thus, the proposed method can be a feasible solution to providing a higher data rate of broadcast and multicast services in the cellular networks. In future work, it will be interesting to investigate the performance of the proposed and conventional methods using a practical channel-estimation method.

Acknowledgment

The authors would like to thank Prof. I. Song at the Korea Advanced Institute of Science and Technology, Daejeon, Korea, for his valuable comments and suggestions that have improved the clarity of this paper.

References

[1] P. Agashe, R. Rezaiifar, and P. Bender, "CDMA2000 high rate broadcast packet data air interface design," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 83–89, Feb. 2004.

[2] M. Chuah, T. Hu, and W. Luo, "UMTS release 99/4 airlink enhancement for supporting MBMS services," in *Proc. IEEE ICC*, Paris, France, Jun. 2004, vol. 6, pp. 3231–3235.

[3] S. Parkvall, E. Englund, M. Lundevall, and J. Torsner, "Evolving 3G mobile systems: Broadband and broadcast services in WCDMA," *IEEE Commun. Mag.*, vol. 44, no. 2, pp. 30–36, Feb. 2006.

[4] M. Luby, T. Gasiba, T. Sockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 235–246, Mar. 2007.

[5] *Physical layer aspects for evolved UTRA*, Feb. 2006. 3GPP, TR 25.814, v.1.2.1.

[6] *Detailed description of the enhanced BCMCS transmit waveform description*, Aug. 2004. 3GPP2, C30-20040823-060.

[7] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[8] M. Inoue and M. Nakagawa, "Space time transmit site diversity for OFDM multi base station system," in *Proc. Int. Workshop Mobile Wireless Commun. Netw.*, Stockholm, Sweden, Sep. 2002, pp. 30–34.

[9] S. K. Kang, Y. H. Kim, and J. Y. Ahn, "Macro diversity coding for broadcast and multicast services in a coded OFDM cellular system," in *Proc. IEEE VTC*, Melbourne, Australia, May 2006, vol. 4, pp. 1590–1594.

[10] *Multiplexing and channel coding (FDD) (Release 6)*, Jun. 2005. 3GPP, TR 25.212, v.6.5.0.

[11] E. K. Hall and S. G. Wilson, "Design and analysis of turbo codes on Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 160–174, Feb. 1998.

[12] D. Divsalar and M. J. McEliece, "Effective free distance of turbo codes," *Electron. Lett.*, vol. 32, no. 5, pp. 445–446, Feb. 1996.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.

[14] *Guidelines for evaluation of radio transmission technologies for IMT-2000*, 1997. Recommendation ITU-R M.1225.

Optimizing Truncated ARQ Scheme Over Wireless Fading Channels

Sangheon Pack, *Member, IEEE*, Xuemin Shen, *Senior Member, IEEE*, and Jon W. Mark, *Life Fellow, IEEE*

***Abstract*—In this paper, we introduce a fractional retransmission (FR) scheme to optimize the performance of a truncated automatic repeat request (ARQ) error control mechanism for transmission over wireless fading channels. For data and real-time traffic packets, we derive the explicit expressions of the average packet loss rate and find the optimal retransmission limit. Simulation results demonstrate that the FR scheme can minimize the packet loss rate.**

***Index Terms*—Automatic repeat request (ARQ), fading channel, fractional retransmission (FR), $M/M/1/K$, optimal retransmission limit.**

I. Introduction

Automatic repeat request (ARQ) is a well-known technique for reliable data transmission, which is particularly preferable for wireless link with burst transmission errors. In ARQ, a sender retries a failed transmission until the transmission is successful or up to a predefined retransmission limit. Unlike forward error correction, ARQ with an infinite retransmission limit cannot bound the transmission latency, which is an important constraint for real-time (RT) traffic transmissions. Therefore, a finite number of retransmissions is typically employed for ARQ, and it is referred to as a truncated ARQ scheme.

The retransmission limit M has a significant impact on the performance of the truncated ARQ scheme. If M is large, the packet loss rate

Manuscript received May 7, 2006; revised June 3, 2007 and July 18, 2007. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Strategic Grant STPGP 257682 and in part by the Ministry of Information and Communication/Institute for Information Technology Advancement under the IT R&D Program (2007-F-038-01, Fundamental Technologies for the Future Internet). The review of this paper was coordinated by Prof. X.-G. Xia.

S. Pack is with the School of Electrical Engineering, Korea University, Seoul 136-701, Korea (e-mail: shpack@korea.ac.kr).

X. Shen and J. W. Mark are with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: xshen@bbcr.uwaterloo.ca; jwmark@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2007.907082

due to transmission errors can be reduced. However, a large value of M increases the waiting time for packet transmissions, and therefore, it can lead to higher buffer overflow. On the other hand, for a small value of M, the possibility of buffer overflow is not significant, but more packets will be lost due to transmission errors. Consequently, it is important to seek an optimal retransmission limit to minimize the total packet loss rate due to both transmission errors and buffer overflows.

In the literature, there have been some research works on the analysis of the optimal retransmission limit in the truncated ARQ scheme and its adaptive application. Vacirca *et al.* [1] investigated the optimal choice for ARQ to improve the transmission control protocol (TCP) performance over wireless links. Shen *et al.* [2] analyzed the effect of the retransmission limit on the TCP-friendly rate control performance over wireless links. Li and Schaar [3] proposed a retransmission limit adaptation scheme for adaptive quality of service provisioning to layered coded video in wireless local area networks. Issariyakul and Hossain [4] classified ARQ schemes and analyzed their performance in multihop wireless networks. Chatterjee *et al.* [5] studied the optimal medium access control (MAC) layer retransmission limit in code division multiple access systems employing both MAC layer retransmission and radio link protocol.

All the aforementioned works use an integer value as an optimal retransmission limit. A concept of *fractional threshold* has been introduced for location management [6] and call admission control [7]. Fractional threshold uses a real number to achieve the optimal performance. In this paper, we propose a fractional retransmission (FR) scheme to enhance the packet loss performance in the truncated ARQ scheme over wireless fading channels. In the FR scheme, a sender maintains a retransmission limit as a real number $M^* = n + \alpha$, where n is an integer, and $0 \leq \alpha < 1$. The sender retransmits an unsuccessful packet up to n times. If the packet also fails at the nth attempt, the sender retransmits the packet with probability α at the $(n+1)$th attempt and drops the packet with probability $1-\alpha$. Simulation results demonstrate that the performance of the truncated ARQ can be optimized by adopting the FR scheme. Our main contributions are two-fold. First, we obtain explicit expressions of the packet loss rates for both data and RT traffic packets and derive the optimal retransmission limit. Second, the proposed FR scheme can minimize the packet loss rate over wireless fading channels with minimal implementation overhead.

The remainder of this paper is organized as follows: Section II describes the system model. By using a discrete-time $M/M/1/K$ queue, the optimal retransmission limit is derived, and the FR scheme is presented in Section III. Simulation results are given in Section IV, followed by concluding remarks in Section V.

II. System Model

We consider a downlink transmission over a wireless Rayleigh fading channel. The base station (BS) has a first-in–first-out buffer with a finite size $B-1$. Packets arrive according to a Poisson distribution with rate λ, which is independent from transmission errors. A time-slotted truncated ARQ scheme is employed for packet transmission, where, at most, one packet can be transmitted during a time slot of length D. If the buffer is empty, an arriving packet is immediately transmitted. On the other hand, if the buffer is full when a packet arrives, the packet is lost due to overflow. If a packet transmission fails at a given time slot, the packet is retransmitted at the next time slot, but the maximum number of retransmissions is bounded by $M-1$. In short, a packet is removed from the BS buffer after being successfully delivered or after M transmission attempts (including the first transmission). Since the feedback message is usually short and well protected, it is assumed error free.

Finite-state Markov processes have been widely used to model Rayleigh fading channels at the packet level because of their mathematical tractability [8]. Therefore, we employ a two-state, i.e., *bad* and *good*, Markov chain to model the wireless fading channel, where time is slotted with a fixed duration D that includes the packet transmission and feedback reception times. In the good state, a packet is successfully transmitted, whereas a packet transmission fails in the bad state. The transition probabilities of these states are given by the following matrix:

$$\mathbf{P} = \begin{pmatrix} p_{bb} & p_{bg} \\ p_{gb} & p_{gg} \end{pmatrix} \tag{1}$$

where p_{xy} is the transition probability from state $x \in \{b, g\}$ to state $y \in \{b, g\}$. Given an average error rate π_b, which is defined as the ratio of the number of erroneous packets to the total number of sent packets, and a burst length l_B, which is defined as the average length of consecutive packet losses, the transition probabilities can be easily computed. The derivation of the fading parameters l_B and π_b from a low level channel error process is beyond the scope of this paper.

III. FR Scheme

In this section, we first derive the optimal retransmission limit for data and RT packets. Subsequently, the FR scheme is proposed.

A. Optimal Retransmission Limit

We use a discrete-time $M/M/1/K$ queue to derive the optimal retransmission limit, where the system size is B. Let λ_B and $\mu_B(M)$ be the arrival and service rates in a time slot when the retransmission limit is $M-1$, respectively. The probability that a packet transmission is successful at the ith transmission attempt $(2 \leq i < M)$ when the channel state at the last transmission of the previous packet is g is given by $p_{gb}p_{bb}^{i-2}p_{bg}$. On the other hand, $p_{bb}^{i-1}p_{bg}$ is the probability that a packet transmission is successful at the ith transmission attempt $(1 \leq i < M)$ when the channel state at the last transmission of the previous packet is b. $p_{gb}{p_{bb}}^{M-2}$ and $p_{bb}{p_{bb}}^{M-2}$ represent the probabilities that the Mth transmission is attempted (either success or failure) when the previous packet transmissions succeeded and failed, respectively. Then, the average service time $1/\mu_B(M)$ (in numbers of time slots) can be computed as

$$\begin{aligned} \frac{1}{\mu_b(M)} &= (1-\pi_b)\Big(p_{gg} + 2p_{gb}p_{bg} + 3p_{gb}p_{bb}p_{bg} + \cdots \\ &\qquad + (M-1)p_{gb}{p_{bb}}^{M-3}p_{bg} + Mp_{gb}{p_{bb}}^{M-2}\Big) \\ &\quad + \pi_b\Big(p_{bg} + 2p_{bb}p_{bg} + 3p_{bb}p_{bb}p_{bg} + \cdots + (M-1) \\ &\qquad \times p_{bb}{p_{bb}}^{M-3}p_{bg} + Mp_{bb}{p_{bb}}^{M-2}\Big) \\ &= (1-\pi_b)\left(1 + \frac{p_{gb}(1-{p_{bb}}^{M-1})}{1-p_{bb}}\right) \\ &\quad + \pi_b\left(\frac{1-{p_{bb}}^{M-1}}{1-p_{bb}} + {p_{bb}}^{M-1}\right) \end{aligned} \tag{2}$$

where the first and second terms on the right-hand side represent the average service times when the previous packet is successfully transmitted and when the previous packet is dropped due to transmission errors, respectively.

The packet loss rate due to buffer overflow is given by [9]

$$O(M) = \frac{(1-\rho_B(M))\,\rho_B(M)^B}{1-\rho_B(M)^{B+1}} \tag{3}$$

where $\rho_B(M) = \lambda_B/\mu_B(M)$. On the other hand, a packet is lost if the channel state remains in the bad state for M time slots. Therefore, the packet loss rate due to transmission error is given by

$$E(M) = \pi_b {p_{bb}}^{M-1}. \quad (4)$$

For a data packet, it can be lost due to either buffer overflow or transmission error. Therefore, the total loss rate of a data packet can be expressed as

$$L(M) = O(M) + (1 - O(M))\, E(M). \quad (5)$$

On the other hand, if the total processing delay (including waiting and service times) of an arrived RT packet exceeds a predefined delay bound, the packet will be useless with respect to applications and discarded. Let q_n be the probability that an arriving packet finds n packets in the system, i.e.,

$$q_n = \frac{1}{1 - O(M)} \left(\frac{(1 - \rho_B(M))\, \rho_B(M)^n}{1 - \rho_B(M)^{B+1}} \right). \quad (6)$$

Let T be the total processing latency in the $M/M/1/K$ queue. Then, the cumulative distribution function (CDF) of T can be defined as

$$\Pr(T \le t) = q_0 \cdot \left(1 - e^{-\mu(M)t}\right) + \sum_{n=1}^{B-1} q_n \cdot \Pr\left((n+1) \text{ packets are serviced within } t|n \text{ packets}\right) \quad (7)$$

where $1 - e^{-\mu(M)t}$ is the CDF of the arriving packet's service time. In the $M/M/1/K$ queue, the service time follows an exponential distribution with rate $\mu(M)$, so that the service time of $(n+1)$ packets (including the arriving packet) follows an $(n+1)$-stage Erlangian distribution. Therefore, (7) can be rewritten as

$$q_0 \cdot \left(1 - e^{-\mu(M)t}\right) + \sum_{n=1}^{B-1} q_n \cdot \int_0^t \frac{\mu(M)\left(\mu(M)x\right)^n}{n!} e^{-\mu(M)x} dx. \quad (8)$$

After some manipulation, $\Pr(T \le t)$ can be derived as [9]

$$\Pr(T \le t) = q_0 \cdot \left(1 - e^{-\mu(M)t}\right) + \sum_{n=1}^{B-1} q_n \cdot \left(1 - \sum_{i=0}^{n} \frac{\left(\mu(M)t\right)^i e^{-\mu(M)t}}{i!}\right). \quad (9)$$

Let δ be the predefined delay bound for an RT packet, which is determined by the application requirements. Then, the packet loss rate due to delay outage is given by

$$D(M) = \Pr(T > \delta) = 1 - \Pr(T \le \delta). \quad (10)$$

Consequently, the total loss rate of an RT packet can be computed as

$$L(M) = O(M) + (1 - O(M))\, E(M) + (1 - O(M))\,(1 - E(M))\, D(M). \quad (11)$$

Obviously, $O(M)$ and $D(M)$ are increasing functions of M, whereas $E(M)$ is a decreasing function of M. Depending on the selected parameters (system size B, delay bound for an RT packet δ, and channel state transition matrix $\mathbf{P}$), $L(M)$ can be an increasing, decreasing, or convex function. If $L(M)$ is an increasing or decreasing function, the optimal retransmission limit M^* is determined by 1 (i.e., the lower bound) or a maximum feasible value for retransmission (i.e., the upper bound). On the other hand, if $L(M)$ is a convex function of M, the optimal retransmission limit M^* satisfies $(\partial L(M)/\partial M)|_{M=M^*} = 0$. Accordingly, M^* can be obtained from a binary search algorithm shown in Algorithm 1, which is similar to that in [6]. Initially, let $Left = 1$ and $Right = MAX$, which are the left and right endpoints of the search interval $[Left, Right]$. MAX is the maximum number of retransmission. If $(\partial L(M)/\partial M)|_{M=Left}(\partial L(M)/\partial M)|_{M=Right} > 0$, $L(M)$ is an increasing or decreasing function of M, and therefore, M^* is determined by $Left$ (increasing function) or $Right$ (decreasing function). Otherwise, it is evaluated based on whether $\partial L(M)/\partial M$ is less than a sufficiently small value ϵ when M is equal to $K = (Left + Right)/2$. If the evaluation result is positive, M^* is given by K, and this algorithm is terminated. On the other hand, if $(\partial L(M)/\partial M)|_{M=K}$ is larger than ϵ,

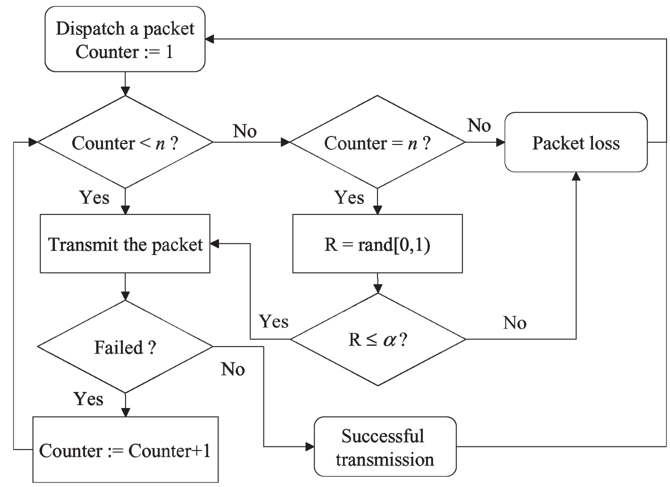

Fig. 1. FR scheme.

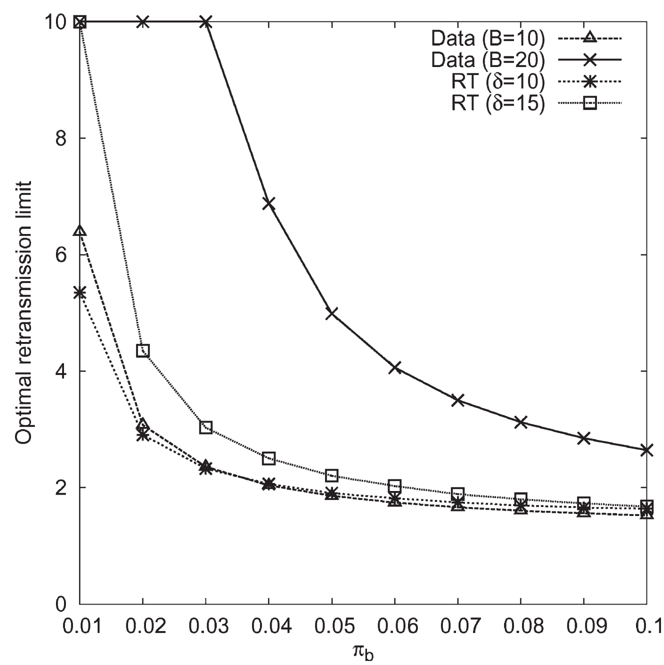

Fig. 2. Optimal retransmission limit ($MAX = 10$).

TABLE I
TOTAL PACKET LOSS RATE: DATA PACKET ($B = 10$, A: ANALYTICAL RESULTS, S: SIMULATION RESULTS)

π_b	M^*	$\rho_B(M^*)$	Fractional (A)	Fractional (S)	Floor (A)	Floor (S)	Ceiling (A)	Ceiling (S)
0.01	6.39	1.08	0.0326977	0.0327	0.0326986	0.03271	0.0326994	0.03272
0.1	1.52	1.06	0.119839	0.1198	0.121144	0.1213	0.120728	0.1209

TABLE II
TOTAL PACKET LOSS RATE: RT PACKET ($B = 10$ AND $\delta = 10$)

π_b	M^*	$\rho_B(M^*)$	Fractional (A)	Fractional (S)	Floor (A)	Floor (S)	Ceiling (A)	Ceiling (S)
0.01	5.35	2.04	0.0848014	0.0848	0.0848019	0.0850	0.084803	0.0849
0.1	1.64	1.07	0.167189	0.1670	0.168221	0.1682	0.167461	0.1677

$(\partial L(M)/\partial M)|_{M=K}(\partial L(M)/\partial M)|_{M=Right}$ is tested. If this value is larger than 0, M^* exists in $[Left, K]$, and thus, $Right$ is set to K; otherwise, M^* exists in $[K, Right]$, and $Left$ is set to K. After that, K is set to $(Left + Right)/2$, and the previous evaluation is repeated until $(\partial L(M)/\partial M)|_{M=K}$ converges to a value less than ϵ.

Algorithm 1 Determination of M^*
$Left \leftarrow 1$; $Right \leftarrow MAX$;
if $(\partial L(M)/\partial M)|_{M=Left}(\partial L(M)/\partial M)|_{M=Right} > 0$ **then**
  if $L(M)|_{M=Left} < L(M)|_{M=Right}$ **then**
    $M^* \leftarrow Left$;
  else
    $M^* \leftarrow Right$;
  end if
end if
$K \leftarrow (Left + Right)/2$;
while $(\partial L(M)/\partial M)|_{M=K} > \epsilon$ **do**
  if $(\partial L(M)/\partial M)|_{M=K}(\partial L(M)/\partial M)|_{M=Right} > 0$ **then**
    $Right \leftarrow K$;
  else
    $Left \leftarrow K$;
  end if
  $K \leftarrow (Left + Right)/2$;
end while
$M^* \leftarrow K$.

B. FR Scheme

Fig. 1 shows the flowchart of the FR scheme. Let $M^* = n + \alpha$, where n is an integer, and $0 \leq \alpha < 1$. If the counter for a packet transmission is less than n, the packet is repeatedly transmitted until it is successfully transmitted. When the counter is equal to n, a random number R is chosen in [0,1). If R is equal to or less than α, the $(n + 1)$th transmission for the packet is conducted; otherwise, the packet transmission fails, and the packet is dropped from the BS buffer. When the packet is successfully transmitted or is dropped, the counter is reset, and the next packet is dispatched.

IV. SIMULATION RESULT

We compare the performance of the FR scheme with floor and ceiling function-based schemes via simulations.[1] In the simulation, the following two channel profiles are used: 1) $\pi_b = 0.01$ and $l_B = 5$ and 2) $\pi_b = 0.1$ and $l_B = 5$. The arrival rate λ_B in a given time slot is 0.8.

Fig. 2 shows the optimal retransmission limit M^* for data and RT packets. In the ARQ scheme, the packet loss rate due to transmission errors can be significantly lowered by means of retransmissions. Hence, reducing buffer overflow has a dominant effect on the total packet loss rate. Consequently, even though π_b increases, only few retransmissions are permitted to keep the buffer overflow at a low level. Therefore, M^* decreases as π_b increases. If B is large, the packet loss due to buffer overflow is not significant, and thus, more transmissions can be made without any concern about buffer overflow. Therefore, for data packets, a larger retransmission limit is allowed when the buffer size is large (i.e., $B = 20$). It can be seen that the optimal retransmission limit is also high when the delay bound δ for RT packets is high. This is because RT packets are less sensitive to transmission latency if δ is large.

As shown in Tables I and II, the gain of the FR scheme is not apparent when the transmission error rate is low (i.e., $\pi_b = 0.01$). This is because most transmission errors can be recovered by a small number of retransmissions in such situation. On the other hand, if the transmission error rate is high (i.e., $\pi_b = 0.1$), it is possible to reduce the total packet loss rate by adopting the FR scheme. In short, the FR scheme can further reduce the total packet loss rate with comparable implementation overhead as the floor and ceiling function-based schemes.

[1] The floor and ceiling functions of a real number x return the largest integer less than or equal to x and the smallest integer not less than x, respectively.

V. CONCLUSION

In this paper, we have derived the optimal retransmission limit for data and RT packets in the truncated ARQ scheme over wireless fading channels. In addition, we have proposed an FR scheme to optimize the performance of the truncated ARQ. Simulation results demonstrate that the FR scheme can reduce the packet loss rates for both data and RT packets compared with integer-value-based adaptive schemes.

REFERENCES

[1] F. Vacirca, A. Vendictis, and A. Baiocchi, "Optimal design of hybrid FEC/ARQ schemes for TCP over wireless links with Rayleigh fading," *IEEE Trans. Mobile Comput.*, vol. 5, no. 4, pp. 289–302, Apr. 2006.
[2] H. Shen, L. Cai, and X. Shen, "Performance analysis of TFRC over wireless link with truncated link-level ARQ," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1479–1487, Jun. 2006.
[3] Q. Li and M. Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 278–290, Apr. 2004.
[4] T. Issariyakul and E. Hossain, "Performance modeling and analysis of a class of ARQ protocols in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3460–3468, Dec. 2006.
[5] M. Chatterjee, G. Mandyam, and S. Das, "MAC layer retransmissions in 1XTREME," in *Proc. IEEE PIMRC*, Sep. 2002, pp. 1452–1456.
[6] Y. Xiao, "Optimal fractional movement-based scheme for PCS location management," *IEEE Commun. Lett.*, vol. 7, no. 2, pp. 67–69, Feb. 2003.
[7] R. Ramjee, D. Towsley, and R. Nagarajan, "On optimal call admission control in cellular networks," *Wirel. Netw.*, vol. 3, no. 1, pp. 29–41, Mar. 1997.
[8] M. Zorzi, R. Rao, and L. Milstein, "On the accuracy of a first-order Markov model for data transmission on fading channels," in *Proc. IEEE ICUPC*, Nov. 1995, pp. 211–215.
[9] D. Gross and C. Harris, *Fundamentals of Queuing Theory*. Hoboken, NJ: Wiley, 1985.