Analytical Modeling for Delay-Sensitive Video Over WLAN

Hossein Bobarshad, *Member, IEEE*, Mihaela van der Schaar, *Fellow, IEEE*, A. Hamid Aghvami, *Fellow, IEEE*, Reza S. Dilmaghani, *Member, IEEE*, and Mohammad R. Shikh-Bahaei, *Senior Member, IEEE*

***Abstract*—Delay-sensitive video transmission over IEEE 802.11 wireless local area networks (WLANs) is analyzed in a cross-layer optimization framework. The effect of delay constraint on the quality of received packets is studied by analyzing "expired-time packet discard rate". Three analytical models are examined and it is shown that M/M/1 model is quite an adequate model for analyzing delay-limited applications such as live video transmission over WLAN. The optimal MAC retry limit corresponding to the minimum "total packet loss rate" is derived by exploiting both mathematical analysis and NS-2 simulations.**

We have shown that there is an interaction between "packet overflow drop" and "expired-time packet discard" processes in the queue. Subsequently, by introducing the concept of *virtual buffer size*, we will obtain the optimal buffer size in order to avoid "packet overflow drop". We finally introduced a simple and yet effective real-time algorithm for retry-limit adaptation over IEEE 802.11 MAC in order to maintain a loss protection for delay-critical video traffic transmission, and showed that the average link-layer throughput can be improved by using our adaptive scheme.

***Index Terms*—Cross-layer, delay, expired-time, retransmission, video, wireless local area network (WLAN).**

I. Introduction

LIVE video streaming applications, such as High Definition TV content (HDTV), Digital Video Broadcasting (DVB), Mobile TV, Video Chat, and Multimedia Messaging Service (MMS) over the IEEE 802.11 Wireless Local Area Network (WLAN), have recently gained particular attention [1].

However, to deliver satisfactory quality-of-service (QoS) to end-users of such networks, there are numerous challenges, among which the restrictions on network bandwidth, meeting delay bounds for timely delivery of multimedia data, and dealing with hostile wireless environment can be named. Various schemes across the communication network layers have been proposed to tackle these problems and for efficient QoS provisioning in WLAN and other wireless networks [2]–[5]. In order to enhance QoS for video over IP applications in the IEEE 802.11, numerous techniques have been designed and proposed. Also extensive research work has recently dealt with cross-layer strategies for QoS improvement over WLANs [6]. Moreover, new resource allocation methods have been recently proposed for improving quality of video streaming application over 3G, ad-hoc, and heterogeneous networks [7], [8]. In [9], a QoS-aware scheduling algorithm for video delivery based on a cross layer approach has been proposed to achieve high perceived video quality by users while satisfying application delay constraint. A joint packetization and retransmission scheme in a delay-bound setting using a cross-layer optimization approach is developed in [10], where the Application and MAC layers jointly determine the optimal packet sizes and retry-limits. Authors in [10] used a real-time greedy algorithm to find a practical solution to this optimization problem. Authors in [11] illustrate a simple strategy across Application (APP), MAC, and Physical (PHY) layers to improve the quality of multimedia. They did not provide analytical solution for the optimal MAC retry-limit. Authors in [12] proposed a novel adaptive cross-layer protection strategy for enhancing the robustness and efficiency of scalable video transmission by performing trade-off between throughput, reliability, and delay depending on the channel conditions and application requirements. Their analysis is based on fluid queuing model without considering the effect of "packet overflow drop". Using cross-layer optimization across the MAC and Application layers in [13], the optimum retry-limit is obtained based on frame error probability and frame expiry time without considering the effect of "packet overflow drop" due to finite buffer size. In [14], authors address the problem of cross-layer optimization of wireless video in real-time, using a classification-based framework. They have investigated the problem of assigning optimal MAC retry-limits for video transmission under delay constraints by using fluid queue model. Overflow is not considered in their work. An "expired-time packet discard" mechanism with adaptive retry-limit is proposed in [15] to improve video streaming over WLAN. This algorithm sets up a retransmission deadline and adaptively makes decision on discarding or (re)sending a packet. An approximate, Fluid-based analytical model for a wireless link with hybrid automatic repeat request (HARQ) has been proposed in [16], which focuses on the delay distribution and packet discard rate over a wireless link. Recently visual entropy, as a quality measure, is optimized in [17] by employing joint PHY-APP layer methods. Wavelet coding is utilized in [17] for optimal power allocation in downlink. Also

Manuscript received December 08, 2010; revised August 15, 2011; accepted October 03, 2011. Date of publication October 26, 2011; date of current version March 21, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sheila S. Hemami.

H. Bobarshad is with the Engineering Department, Tarbiat Modares University, Tehran, Iran (e-mail: hossein.bobarshad@ieee.org).

M. van der Schaar is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: mihaela@ee.ucla.edu).

A. H. Aghvami, R. S. Dilmaghani, and M. R. Shikh-Bahaei are with the Centre for Telecommunication Research, King's College London, London WC2R 2LS, U.K. (e-mail: hamid.aghvami@kcl.ac.uk; reza.shams_dilmaghani@kcl.ac.uk; m.sbahaei@kcl.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2011.2173477

a new coding scheme, namely Tetrys coding, is used in [18] to improve the video quality in real-time applications by reducing the delay for recovering lost packets at the receiver. To evaluate the performance of different QoS provisioning strategies, various analytical models based on finite and infinite queuing model are employed in the literature. In [19], the Fluid and M/G/1 models have been employed for the analysis and evaluation of its cross layer retry-limit optimization technique for enhancement of video transmission over IEEE 802.11 WLANs. Also a queuing-based analytical model is used in [20] for the playback buffer at the receiver, and the model is employed for deriving mathematical expressions for video quality in terms of delay and packet loss. The focus in their work is on adaptive playback buffer management with no regard to transmit buffer.

The effect of adaptive MAC retry-limit on network congestion in a multiuser environment is analyzed by [21] for delay-sensitive video over WLAN, where joint optimization of MAC and Application layers is studied for video streaming applications over 802.11e wireless networks. Using a low-complexity analytical model, in [22], we derived closed forms for MAC-layer service time and "total packet loss rate" for real-time video over WLAN. In [22], we assumed that packet loss is either due to link layer deficiency or caused by limited buffer size at the transmitter. However, when we deal with delay-limited applications such as real-time video streaming, time expiry is a major source of packet loss. Here we aim to address this issue by providing a new analytical model for delay-sensitive video transmission over WLANs.[1] The contributions in this work can be summarized as follows.

We will derive mathematical expressions for the "expired-time packet loss rate" and the "total packet loss rate". These parameters will be incorporated in optimizing live video transmission in a cross-layer platform. Then we will show that the "expired-time packet discard" mechanism can be modeled by a second "packet overflow drop" process. Furthermore, we analytically derive the optimum size for the physical buffer in order to avoid actual "packet overflow drop". Finally, for optimizing transmission against "expired-time packet discard" mechanism, we propose a novel real-time retry-limit adaptation algorithm in this paper. In particular, we will examine two scenarios: 1) with "expired-time packet discard" mechanism, but *without* "packet overflow drop" due to assuming a very long buffer size, and 2) with concurrent "expired-time packet discard" *and* "packet overflow drop" mechanisms in the queue. We will show that there is an interaction between the two packet drop mechanisms.

The rest of this paper is organized as follows. In Section II, the scenario and the queuing system model considered in this paper are discussed. Subsequently, three sources of packet loss according to this scenario are briefly explained. In Section III, the "expired-time packet discard rate" is obtained by applying the M/M/1/K, M/G/1, and M/M/1 models. Then, the achieved results from this section are used to select the most efficient mathematical model.

In Section IV, we analyze our scenario for the case where there are both "packet overflow drop" and "expired-time packet discard" mechanisms in the queue. In the proposed analysis, the trade-off between two packet drop mechanisms in the queue, the concept of virtual buffer length, and evaluation of the optimum buffer length will be investigated. In Section V, M/M/1 model is exploited to solve the retry-limit optimization problem and the mathematical conditions for validity of our retry-limit optimization problem will be explored. Also the simulation and analysis of the packet delay performance will be studied. In Section VI, a new retry-limit adaptation algorithm is proposed based on the M/M/1 model formulations. Finally, conclusions are drawn in Section VII.

[1]The cross-layer structure and interconnections across layers and also the assumptions about packet overflow in this work are the same as in [19] and [22].

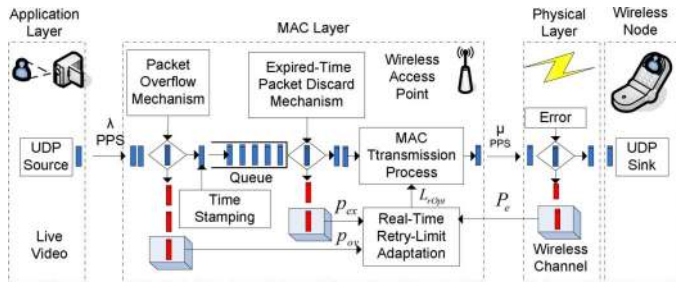

Fig. 1. Live video communication over a wireless network.

TABLE I
SUMMARY OF NOTATIONS AND ABBREVIATIONS

Notati	Description	Notatio	Description
λ	Incoming Average Rate	p_T	Total Packet Loss Rate
μ	Average Service Rate	p_L	Packet Link Loss Rate
T_{ex}	Packet Expiry Time	p_{ov}	Packet Overflow Drop Rate
P_e	Packet Error Rate (PER)	K_A	Hypothetical Finite Buffer Size
$a(x)$	Input traffic distribution	p_{ex}	Expired-Time Packet Discard Rate
$b(x)$	Service Time Distribution	T_q	Waiting Time in Queue
L_r	Retry-Limit	$W(x)$	Cumulative Distribution Function (CDF) Of "Waiting Time In The
K	Simulation Buffer Size	p_{TQ}	Packet Drop Rate At The Queue

II. System Model

We consider the downlink transmission scenario from access point (AP) to station shown in Fig. 1, in which the live video packets are transmitted according to the IEEE 802.11 WLAN settings. Notations and abbreviations in this paper are summarized in Table I. As in our previous work [22], we model this scenario by exploiting a queuing system. The live video source generated by camera is modeled by traffic distribution $a(x)$ with average rate of λ packets per second (PPS), and the wireless access point is viewed as a queuing system with service time distribution $b(x)$ and with an average service rate of μ PPS. In fact, $b(x)$ models the IEEE 802.11 MAC block transmission process. We assume that the other delay times from the video camera to the wireless queue are negligible in comparison with the waiting

time in the queue. Packet loss due to the wireless time-varying channel and due to the collisions—with packets sent by other access points and nodes—is considered through the packet error rate (PER) "P_e" parameter. The access point adaptively adjusts the retry-limit frame by frame. In the retry-limit adaptation block, the optimum retry-limit (L_r) is obtained in order to minimize the "total packet loss rate" (p_T), by virtue of the fact that increasing retry-limit (L_r) will decrease "packet link loss rate" (P_L) and increases both overflow and expired-time packet drop rates. In the above scenario, each packet may be lost either due to drop from the queue at the wireless access point or due to channel errors at the wireless link. Packet dropping from the queue, in turn, can take place when the number of packets in the queue exceeds the buffer length or some packets become expired. Here we briefly explain the above sources of packet loss.

A. Packet Loss Rate and Packet Overflow Drop Rate

The probability of packet loss over the wireless channel after L_r packet retransmissions by the wireless access point, namely the "*packet link loss rate*" (P_L), is given by

$$p_L = P_e^{L_r+1}. \tag{1}$$

In our analysis in [22], we showed the "packet overflow drop rate" $(P_{ov_{(M)}})$[2] in M/M/1 model is given as

$$p_{ov_{(M)}} = \rho^{K_A+1} = \left(\rho_0 \frac{1-P_e^{L_r+1}}{1-P_e}\right)^{K_A+1} \tag{2}$$

where $\rho_0 = \lambda/\mu_0$ and K_A is hypothetical finite buffer size.

B. Expired-Time Packet Discard Rate

We assume that each packet, upon arrival in the queue, is stamped with an arrival time. The difference between this stamped time and current time indicates the packet waiting time in the queue. Should the packet waiting time become greater than a fixed expiry time (T_{ex}), the packet is discarded. This process is known as "*obsolete packet drop*" or "*expired-time packet discard*".

Packet expiry time (T_{ex}) is the deadline for a video packet to receive at the video encoder of the mobile node while it is still valid in the current video frame. According to [15], if we assume a real-time video sequence with group of pictures (GOP) of size α and an inter-frame interval τ_f, the expiry time for the kth video packet in the jth frame within the jth GOP, $P_{i,j}^{(k)}$, can be calculated as

$$T_{ex} = ((i-1)\alpha + (j-1))\tau_f + R(P_{i,j}^{(k)}) \tag{3}$$

where $R(P_{i,j}^{(k)}) = \tau_f(M(F_{i,j})+1)$ and $M(F_{i,j})$ is the number of frames inter-coded with frame $F_{i,j}$ (the jth frame within the ith GOP). The "expired-time packet discard rate", p_{ex}, can be defined as the probability for an arriving packet to wait more than T_{ex} in the queue. Throughout this work, the methodology in [24] is incorporated for scheduling the packets given transmission delay. Denoting the random variable T_q as the "waiting time" in queuing theory, we have

$$p_{ex} = \Pr\{T_q > T_{ex}\}. \tag{4}$$

The cumulative distribution function (CDF) of "waiting time in the queue", $W(x)$, is the probability that T_q is less than or equal to x. That is

$$W(x) = \Pr\{T_q \le x\}. \tag{5}$$

Hence, from (4) and (5), we have

$$p_{ex} = 1 - W(T_{ex}). \tag{6}$$

C. Total Packet Loss Rate

The above three packet loss rates lead to the "total packet loss rate", (p_T):

$$p_T = p_{TQ} + (1-p_{TQ}) \cdot p_L \tag{7}$$

where p_L is obtained from (1) and p_{TQ} is the packet drop rate at the queue which depends on the "packet overflow drop rate" (p_{ov}) and the "expired-time packet discard rate" (p_{ex}):

$$p_{TQ} = p_{ov} + (1-p_{ov})p_{ex}. \tag{8}$$

III. Deriving the "Expired-Time Packet Discard Rate"

In this section with assuming that there is no "packet overflow drop", our aim is to find the "expired-time packet discard rate". In the following, we first apply the finite-queue M/M/1/K model which is commonly used in the analysis of the IEEE 802.11 WLAN systems. Subsequently, we employ the M/G/1 and M/M/1 models in order to calculate the "expired-time packet discard rate". Finally, the accuracy of these models is verified by exploiting Network Simulation 2 (NS-2) [25] results.

A. M/M/1/K Model-Based Analysis

From (6), we can see that in order to obtain p_{ex}, we need to calculate waiting time PDF at $x = T_{ex}$, $W_(k)(T_{ex})$. We use the expression for waiting time in M/M/1/K [23, Ch. 2]:

$$W_{(K)}(x) = 1 - \sum_{n=1}^{K-1} q_n \sum_{i=0}^{n-1} \frac{(\mu x)^i e^{-\mu x}}{i} \tag{9}$$

where q_n is defined as

$$q_n = \frac{p_n}{1-p_K} \quad \text{and} \quad p_n = \frac{(1-\rho)\rho^n}{1-\rho^{n+1}} \tag{10}$$

where $\rho = \lambda/\mu$ and μ is obtained from

$$\mu = \frac{\mu_0}{r} \quad \text{and} \quad r = \frac{1-P_e^{L_r+1}}{1-P_e} \tag{11}$$

[2]Note that the index (M) shows that the corresponding variable is associated with the M/M/1 model. In the rest of this paper, we use index (K) for the variables in M/M/1/K model, (G) for the variables in M/G/1 model, and (S) for NS-2 simulation results.

where P_e is the packet error rate at the link layer. From (6) and (9), the "expired-time packet discard rate" in M/M/1/K can be obtained as

$$p_{ex_{(K)}} = \sum_{n=1}^{K-1} q_n \sum_{i=0}^{n-1} \frac{(\mu T_{ex})^i e^{-\mu T_{ex}}}{i!}. \quad (12)$$

B. M/G/1 Model-Based Analysis

Similar to our discussion in [22] for (approximately) Poisson-distributed packet arrival processes, M/G/1 model can be employed with Laplace transform of the waiting time PDF, $W^*_{(G)}(s)$, can be obtained from [26, eq. 5.105]

$$W^*_{(G)}(s) = \frac{(1-\rho)s}{s - \lambda - \lambda B^*(s)} \quad (13)$$

where $B^*(s)$ is the Laplace transform of $b(x)$, which can be obtained from [22, eq. 18] as

$$B^*(s) = L[b(x)] = \frac{\mu_0}{s+\mu_0} + \sum_{n_r=1}^{L_r} P_e^{n_r} \left(\frac{\frac{\mu_0}{n_r+1}}{s + \frac{\mu_0}{n_r+1}} - \frac{\frac{\mu_0}{n_r}}{s + \frac{\mu_0}{n_r}} \right). \quad (14)$$

From (6), the "expired-time packet discard rate" in the M/G/1 model will be

$$p_{ex_{(G)}} = 1 - W_{(G)}(T_{ex}) \quad (15)$$

where $W_{(G)}(T_{ex})$ is the inverse Laplace transform of (13) at $x = T_{ex}$. The detail of calculation of $W_{(G)}(T_{ex})$ from (13) is provided in Appendix A.

C. M/M/1 Model-Based Analysis

According to [26, eq. 5.123], the packet waiting time CDF in the queue for M/M/1 model is

$$W(x) = 1 - \rho e^{-(\mu-\lambda)x} (x \geq 0) \quad (16)$$

where $\rho = \lambda/\mu$. Using [22, eq. 26], and substituting $W(x)$ (with $x = T_{ex}$) from (16) into (6), the "expired-time packet discard rate" in M/M/1 model will be

$$p_{ex_{(M)}} = \rho e^{-(\mu-\lambda)T_{ex}} = \left(\rho_0 \frac{1 - P_e^{L_r+1}}{1 - P_e} \right) e^{-\left(\frac{\mu_0(1-P_e)}{1-P_e^{L_r+1}} - \lambda \right) T_{ex}} \quad (17)$$

where $\rho_0 = \lambda/\mu_0$.

D. Simulation and Evaluation of the Models

The NS-2 network simulator is used for simulation of the "expired-time packet discard" process in the scenario which is illustrated in Fig. 1. The simulation parameters are set as in Table II. The access point forwards the traffic towards the wireless node via a typical IEEE 802.11b wireless fading channel using a two-ray propagation model. Simulation of the "expired-time packet discard" process in NS-2 environment required some changes in the NS-2's C++ source code. In particular, a time stamp on the header of every packet upon arrival into the queue was augmented. Then, during every en-queuing and de-queuing routine, the entire packets in the queue are checked and those packets which have spent more than T_{ex} seconds in the queue are discarded. The "expired-time packet discard rate" versus MAC retry-limit according to the M/M/1/K, M/G/1, and M/M/1 models, and also based on the NS-2 simulation, are demonstrated in Fig. 2 with the parameters set out in Table II. The service rate in the IEEE 802.11b MAC layer simulation is dependent on routing, transmission power, wireless physical bandwidth, and packet data rate. Therefore, in order to compare the analytical models with simulation results under similar conditions, we have set the parameters using a similar method to our approach in [22, Section VI-A]. In Table II, the service rate of the wireless link in the simulation, $\mu_{0_{(S)}}$, is measured under no retransmission.

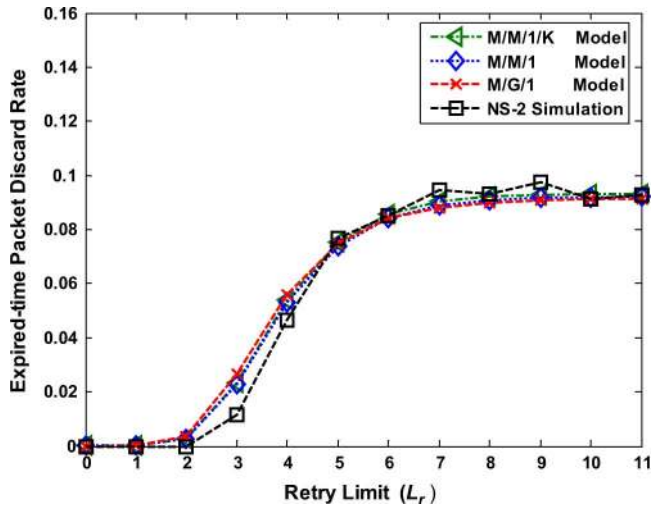

Fig. 2. Comparing the M/M/1/K, M/M/1, and M/G/1 models to NS-2 simulation results (parameters are set according to Table II).

Dash in Table II shows that the parameter is not defined for the respective model. According to Fig. 2 and in line with our discussion in [22], we will employ M/M/1 model here for analysis of the system with both "packet overflow drop" and "expired-time packet discard" mechanisms.

IV. Analyzing System With "Packet Overflow Drop and Expired-Time Packet Discard" Mechanism

Previously, we studied the case where there was only one of the "packet overflow drop" or the "expired-time packet discard" mechanisms present in the queue, respectively, in [22] and in the above sections. Here, we will consider a packet transmission scheme with both packet droppings. From the previous sections, we observed that the M/M/1 queuing model is not only a simple but quite an accurate model for analyzing IEEE 802.11 real-time

TABLE II
PARAMETER VALUES FOR THE ANALYTICAL MODELS AND NS-2 SIMULATION RESULTS

Parameter \ Model	M/M/1/K	M/M/1	M/G/1	NS-2 Simulation
λ (PPS)	260	260	260	260
$\mu_{0_{(Model)}}$ (PPS)	$\mu_{0_{(K)}}$ 453.6	$\mu_{0_{(M)}}$ 453.6	$\mu_{0_{(G)}}$ 462	$\mu_{0_{(S)}}$ 472
T_{ex} (Seconds)	0.2	0.2	0.2	0.2
BasicRate	-	-	-	2M
RTS Threshold	-	-	-	500
ShortRetryLimit	-	-	-	7
(L_r+1) : LongRetryLimit	1-12	1-12	1-12	1-12
Buffer Length	50	-	-	50
Packet Length	-	-	-	1024
P_e PER	0.4	0.4	0.4	0.4

video transmission. Hence, the M/M/1 model is our reference model in this section.

A. Derivation of "Packet Overflow Drop Rate" and "Expired-Time Packet Discard Rate"

To calculate "expired-time packet discard" rate in the presence of "packet overflow drop" process, we need to find the probability that a packet is not dropped due to overflow and can find a place in the queue, but is discarded by expired-time mechanism. We define this "expired-time packet discard rate" by p'_{ex} which can be obtained as

$$p'_{ex} = (1 - p_{ov}) \cdot p_{ex} \tag{18}$$

where p_{ov} and p_{ex} are from (2) and (17), respectively. Hence, we have

$$p'_{ex} = (1 - \rho^{K_A+1})\rho e^{-(\mu-\lambda)T_{ex}}. \tag{19}$$

The definition of the "packet overflow drop rate" when there is "expired-time packet discard" mechanism in the queue (p'_{ov}) can be written as

$$p'_{ov} = \Pr\{N - N_{ex} \geq K_A + 1\} \tag{20}$$

where N is the total number of packets entered the queuing system and N_{ex} is the total number of packets which are expired and discarded from the queue.

Invoking the definition of p_{ex} as

$$p_{ex} = \frac{N_{ex}}{N}. \tag{21}$$

Equation (20) can be rewritten as

$$p'_{ov} = \Pr\left\{N \geq \frac{K_A + 1}{1 - p_{ex}}\right\} = \rho^{\left(\frac{K_A+1}{1-p_{ex}}\right)}. \tag{22}$$

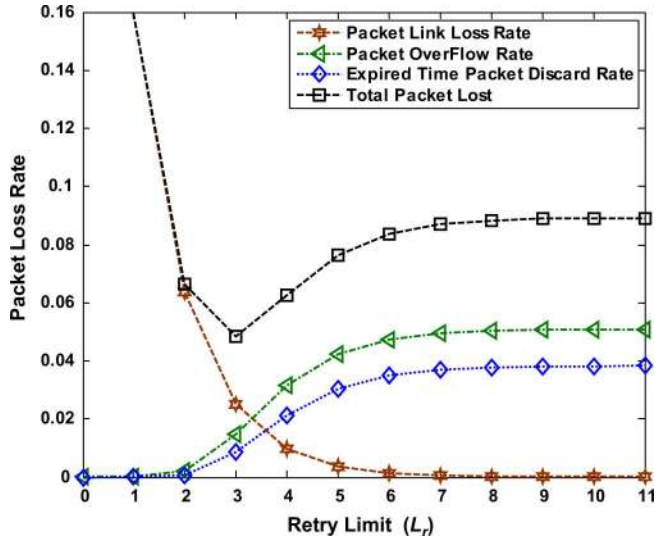

Fig. 3. "Packet link loss" (1), "packet overflow drop rate" (23), "expired-time packet discard rate" (19), and their summation (24) versus MAC retry limit, based on M/M/1 queuing model ($\lambda = 260$ PPS, $\mu_0 = 455.8$ PPS, $P_e = 0.4$, $K_A = 50$, and $T_{ex} = 0.21$ s).

Substituting p_{ex} from (17), we have

$$p'_{ov} = \rho^{\left(\frac{K_A+1}{1-\rho e^{-(\mu-\lambda)T_{ex}}}\right)}. \tag{23}$$

Fig. 3 shows the "packet link loss rate" given in (1), the "packet overflow drop rate" according to (23), and the "expired-time packet discard rate" (19). The "total packet loss rate", p'_T, can be written as

$$p'_T = p'_{TQ} + (1 - p'_{TQ})p_L \tag{24}$$

where p_L is derived from (1) and p'_{TQ} is the "total packet drop rate" in the queue which is given by

$$p'_{TQ} = p'_{ov} + p'_{ex} \tag{25}$$

where p'_{ex} and p'_{ov} are obtained from (19) and (23), respectively. From Fig. 3, it can be seen that there is a minimum point for p'_T that takes place around $L_r = 3$, which can be considered as the optimal value for retry-limit L_r. We have conducted NS-2 simulations for the case that there are both "packet overflow drop" and "expired-time packet discard" mechanisms in the queue. In the simulations, the buffer length is set at $K = 50$ and we assumed an expiry time of $T_{ex} = 0.21$ s. The other parameters are set as in Table II. The NS-2 simulation-based results are given in Fig. 4 for link loss rate $(p_{L_{(S)}})$, "packet overflow drop rate" $(p_{ov_{(S)}})$, "expired-time packet discard rate" $(p_{ex_{(S)}})$, and the sum of the three packet loss rates $(p_{T_{(S)}})$ in the case where there are both packet dropping mechanisms in the queue.

B. Trade-Off Between "Packet Overflow Drop Rate" and "Expired-Time Packet Discard Rate"

Fig. 5(a) and (b) is plotted in the same way as in Figs. 3 and 4 but with different values of expiry time, $T_{ex} = 0.19$ s. Comparing the above results with different values of expiry time

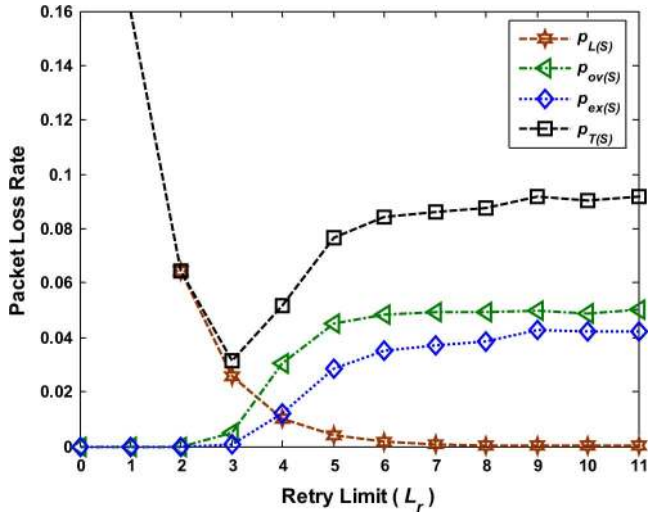

Fig. 4. "Packet link loss rate" ($P_{L_{(S)}}$), "packet overflow drop rate" ($p_{ov_{(s)}}$), "expired-time packet discard rate" ($p_{ex_{(s)}}$), and their summation ($p_{T_{(s)}}$) versus MAC retry limit, made by NS-2 simulation (NS-2 parameters are set as in Table II).

(given that $p'_{ex} > 0$ and $p'_{ov} > 0$) leads to the following properties:

$$\begin{cases} p'_{ov} < p'_{ex} \rightarrow T_{ex} < T'_{ex} \\ p'_{ov} = p'_{ex} \rightarrow T_{ex} < T'_{ex} \\ p'_{ov} > p'_{ex} \rightarrow T_{ex} < T'_{ex}. \end{cases} \tag{26}$$

In (26), T'_{ex} is defined as a certain value of expiry-time where $p'_{ov} = p'_{ex}$. The above properties show that there is a trade-off between p'_{ex} and p'_{ov}. That is, increasing p'_{ex} will imply a decrease in p'_{ov} and vice versa. The simulation results in Fig. 6 clearly show the above properties. For a fixed retry-limit ($L_r = 4$), we ran simulations for 34 different values of expiry time ($T_{ex} = 0.001 - 0.3$ s). According to Fig. 6, p'_{ov} and p'_{ex} curves cross at the point where $T_{ex} = T'_{ex} \cong 0.192$.

In the following section, we calculate this special value of the expiry-time, at which the "packet overflow drop rate" and "expired-time packet discard rate" are equal.

C. Expired-Time Packet Dropping Mechanism and Virtual Buffer Size Concept

In the previous section, we defined T'_{ex} as the expiry-time where the "packet overflow drop rate" and "expired-time packet discard rate" are equal; hence, from (2) and (17), we can write

$$\left(\rho_0 \frac{1 - P_e^{L_r+1}}{1 - P_e}\right)^{K_A+1} = \left(\rho_0 \frac{1 - P_e^{L_r+1}}{1 - P_e}\right) e^{-\left(\frac{\mu_0(1-P_e)}{1-P_e^{L_r+1}} - \lambda\right) T'_{ex}}. \tag{27}$$

Using (27) and rewriting T'_{ex}, we have

$$T'_{ex} = \alpha \frac{K_A}{\lambda} \text{ where } \alpha = \frac{\ln\left(\rho_0 \frac{1-P_e^{L_r+1}}{1-P_e}\right)}{1 - \left(\rho_0 \frac{1-P_e^{L_r+1}}{1-P_e}\right)^{-1}}. \tag{28}$$

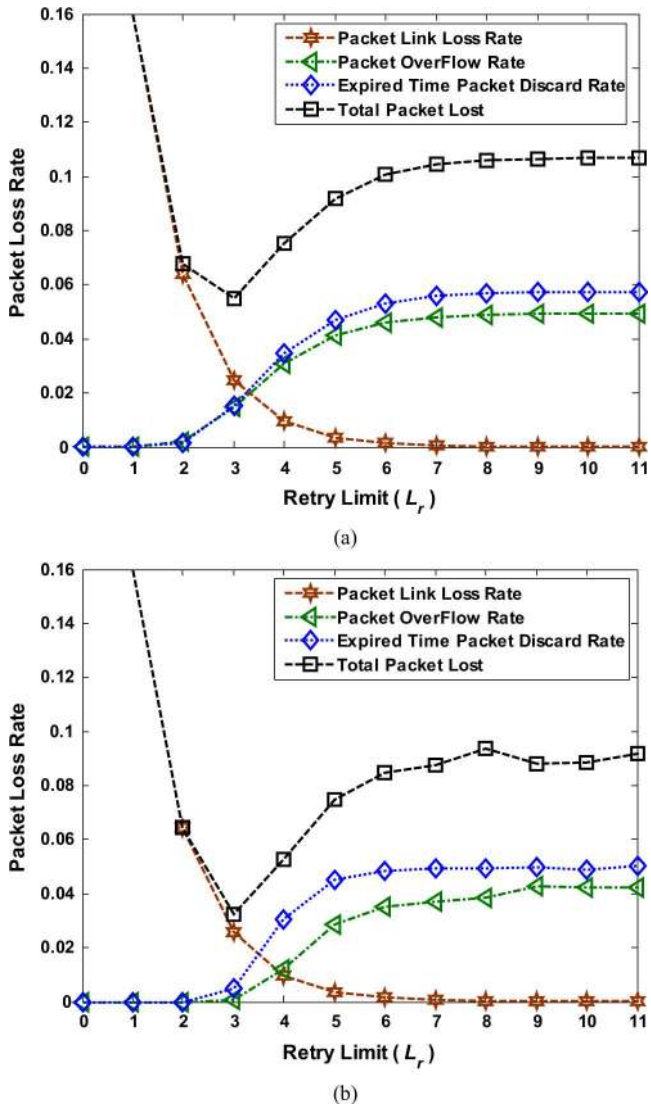

Fig. 5. "Packet link loss rate", "packet overflow drop rate", "expired-time packet discard rate", and their summation versus MAC retry limit, made by: (a) M/M/1 model and (b) NS-2 simulation (parameters are set as in Table II and $T_{ex} = 0.185$ s).

The value of α is close to one; hence, we have

$$T'_{ex} \cong \frac{K_A}{\lambda}. \tag{29}$$

For example for values of $\lambda = 260$ PPS, $\mu_0 = 455.8$ PPS, $P_e = 0.4$, $L_r = 4$, and $K_A = 50$, the value of α is 0.97 and T'_{ex} from (28) and (29) is 0.187 and 0.1923, respectively. Now we look at this from another perspective. Assuming T_q as the mean value of waiting time in the queue, from Little's low in queuing theory, we have

$$T_q = \frac{K_A}{\lambda}. \tag{30}$$

Comparing (29) and (30) implies that having $T'_{ex} \cong T_q = K_A/\lambda$ will guarantee that the "packet overflow drop rate" and "expired-time packet discard rate" are equal. That is, should the

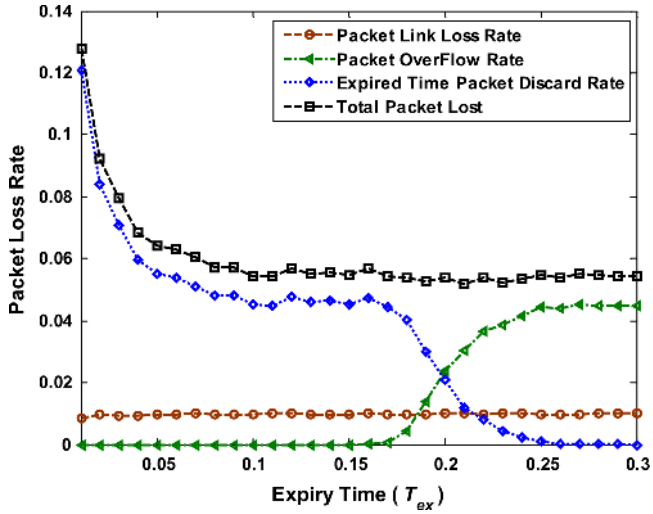

Fig. 6. Simulation results for $(p_{L_{(s)}})$, $(p_{ov_{(s)}})$, $(p_{ex_{(s)}})$, and their summation $(p_{T_{(s)}})$ versus expiry-time (p_{ex}) (NS-2 parameters are set as in Table II where $L_v = 4$).

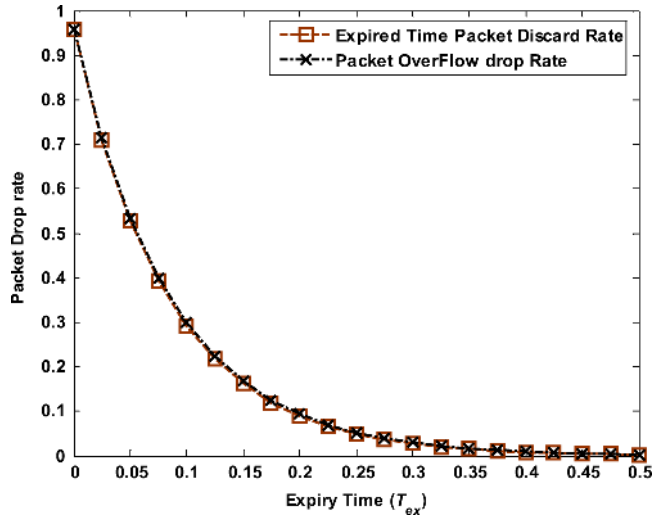

Fig. 7. "Packet overflow drop rate" from (31) and "expired-time packet discard rate" due to (17) versus T_{ex} in M/M/1 queuing analysis model.

expiry-time equal the mean value of waiting time in the queue, the values of "packet overflow drop rate" and "expired-time packet discard rate" will be equal. To verify this result in (2), buffer length is replaced with $K_A = T'_{ex} \cdot \lambda$ from (29). Hence, we have

$$p_{ov} = \left(\rho_0 \frac{1 - P_e^{L_r+1}}{1 - P_e}\right)^{T_{ex}\lambda+1}. \tag{31}$$

We have plotted the "packet overflow drop rate" from (31) and "expired-time packet discard rate" from (17) versus T_{ex} in Fig. 7. By looking at this figure, we can conclude that having "expired-time packet discard" mechanism with expiry time value of T_{ex} is effectively equivalent to having a secondary overflow drop process with a virtual buffer size of $T_{ex}\lambda$. Therefore, a queuing system with both actual overflow drop and "expired-time packet discard" mechanisms is equal to a system with a real buffer size K_A and a virtual buffer size K'_A as

$$K'_A = T_{ex}\lambda. \tag{32}$$

Depending on the values of buffer size (K_A) and virtual buffer size (K'_A), the "total packet drop rate" in the queue, (p'_{TQ}), is divided between p'_{ov} and p'_{ex} where we have $p'_{TQ} = p'_{ov} + p'_{ex}$.

In the next section, we aim to use this property to avoid "packet overflow drop" in the system. According to the above results, any fixed T_{ex} in our model can be attributed to a virtual buffer length. Having a physical buffer of smaller length than the virtual buffer would entail some dropped packets due to overflow, whereas these packets are still valid to be sent given their expiry time, T_{ex}.

Thus, we are trying to determine what should the physical buffer length be such that it can accommodate these packets until they are valid to be sent. The following subsection aims to answer this question by calculating the optimum physical buffer length. Also in Section IV-D, we show that setting the physical buffer length at a higher value than the optimum value would unnecessarily increase the cost.

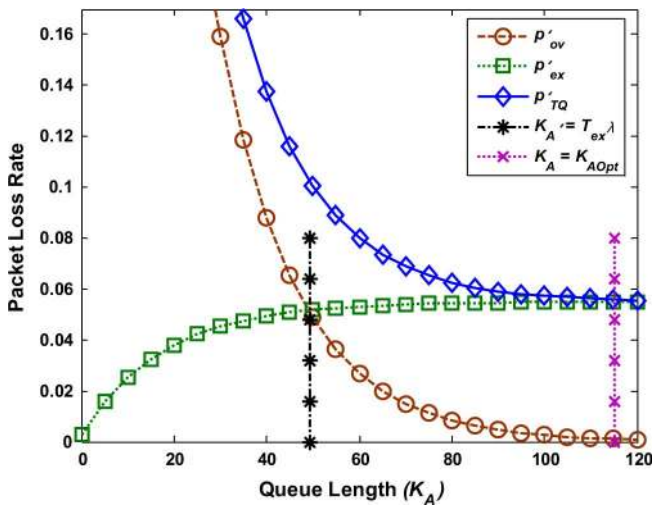

Fig. 8. "Packet overflow drop rate" (p'_{pv}), "expired-time packet discard rate" (p'_{ex}), and their summation versus buffer size (K_A) along with the value of $K'_A = T_{ex}\lambda$ and K_{AOpt} from (34) ($\lambda = 260$ PPS, $\mu_0 = 455.8$ PPS, $P_e = 0.4$, $L_r = 4$, and $T_{ex} = 0.19$ s).

D. Optimum Buffer Size for Avoiding "Packet Overflow Drop"

We showed that "packet overflow drop rate" and "expired-time packet discard rate" can be controlled by K_A and K'_A, respectively. Since the value of K'_A is fixed, we aim to find the optimum value of real system buffer size, K_A, which minimizes the "total packet drop rate" in the queue $(p'_{TQ} = p'_{ov} + p'_{ex})$.

In Fig. 8, we have plotted p'_{ex} from (19), p'_{ov} from (23), and the "total packet drop rate" at queue $(p'_{TQ} = p'_{ov} + p'_{ex})$ versus buffer size (K_A). As we can see from Fig. 8, p'_{TQ} is a descending curve with no finite absolute minimum. Hence, the practically optimum buffer size (K_{AOpt}) can be found through satisfying the following:

$$\frac{dp'_{TQ}}{dK} \leq \gamma \tag{33}$$

where γ is a small constant that can be obtained from Fig. 8 by determining the point above which "total packet drop rate"

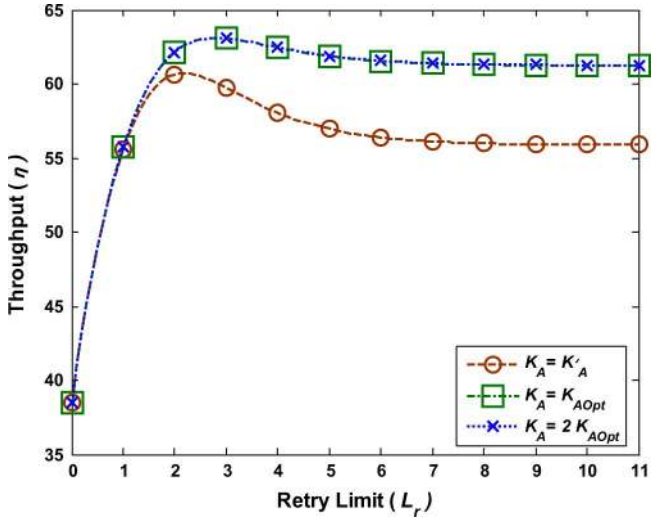

Fig. 9. Comparing average throughput for different values of buffer length based on analytic model results.

in queue (p'_{TQ}) varies insignificantly. Using (19) and (23) in (33), the optimum buffer size, K_{AOpt}, can be obtained from the following recursive formula:

$$K_{AOpt} = \log_\rho \left(\frac{\gamma}{\kappa \rho^{(K_{AOpt}-1)(\psi-1)} - p_{ex}} \right)$$

where

$$\psi = \frac{1}{1 - p_{ex}} \tag{34}$$

where p_{ex} is given by (17).

In Fig. 9, for the parameters set of $\gamma = 0.001$, $\lambda = 260$ PPS, $\mu_0 = 455.8$ PPS, $P_e = 0.4$, $L_r = 4$, and $T_{ex} = 0.19$ s, we have plotted the system throughput when buffer size takes on three different values of $K_A = K'_A$, $K_{A_{Opt}}$, and $2K_{A_{Opt}}$. The system throughput (η) in Fig. 9 is calculated from the following equation:

$$\eta = \lambda(1 - p'_T) \tag{35}$$

where p'_T is obtained from (24).

As it can be seen, throughput has improved by changing the buffer size from K'_A to $K_{A_{Opt}}$, but it is nearly invariant in response to changing from $K_{A_{Opt}}$ to $2K_{A_{Opt}}$. To explain this behavior in Fig. 8, the value of K_{AOpt} is highlighted by a vertical line. At the point where $K_A = K_{AOpt}$, the value of "packet overflow drop rate" is almost zero. Therefore, at this point $p'_{ov} \cong 0$; hence, $p'_{TQ} \cong p_{ex}$ and from (25), the "total packet loss rate" while avoiding "packet overflow drop" p'_{TA} can be written as

$$p_{TA} = p'_T|_{K_A = K_{AOpt}} = p_{ex} + (1 - p_{ex}) \cdot p_L. \tag{36}$$

To verify the throughput curve behavior in Fig. 9, we have used real video traffic simulation in the next section.

E. Real Video Traffic Simulation

The simulation results in previous Sections III-D and VI were based on exponential incoming traffic. In this section, a real video traffic source is used for the simulation in order

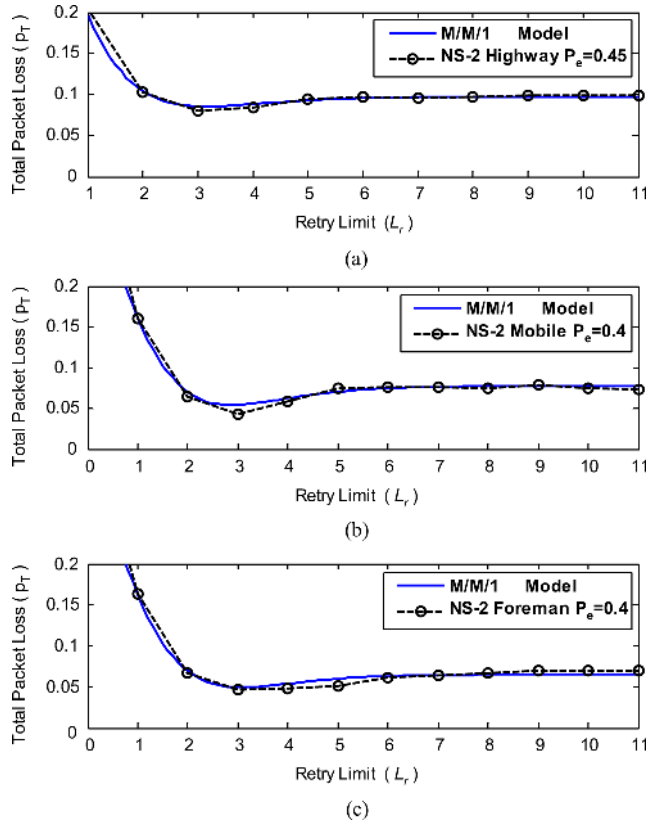

Fig. 10. "Total packet loss rate" for different values of for both real video NS-2 simulation and the M/M/1 model. (a) Highway. (b) Mobile. (c) Foreman.

to confirm the accuracy of our models. The same real video transmission simulation set-up as in our previous work [22] is used. In this work, we have assumed a system performing real-time (live) streaming but not real-time encoding. The encoder performs the encoding offline and stores the packets. Only the streaming (scheduling, etc.) was performed online using the NS-2 simulator. As mentioned in our previous work [22], after compressing the CIF format video sequence using the MPEG4 format (see our previous work for details of the parameter settings), each video frame will be fragmented into the maximum size of 1000 bytes. Furthermore, by adding 20 bytes Internet Protocol (IP) header and 8 bytes UDP header, the maximum packet size will be 1028 bytes in transmission. In real video simulations, Gilbert-Elliott error channel is used. The simulations are run 12 times, each time with a different L_r. In Fig. 10, the "total packet loss rate" (p_T) for real video simulations and for M/M/1 model are plotted. The curve corresponding to M/M/1 model is given along with simulation results for "Highway", "Mobile", and "Foreman" videos as input traffic in Fig. 10(a)–(c), respectively. Also in Fig. 10, from both simulation and the M/M/1 model, it can be seen that "total packet loss rate" varies with L_r, with a minimum point occurring at $L_r \cong 3$. PSNR performance is also shown for different values of retry limit in Fig. 11. According to this figure, our optimal retry limit provides high PSNR quality, too. Due to this observation and the simplicity and accuracy of the M/M/1 model for analyzing IEEE 802.11 live video transmission process, we use this model to calculate a closed-form

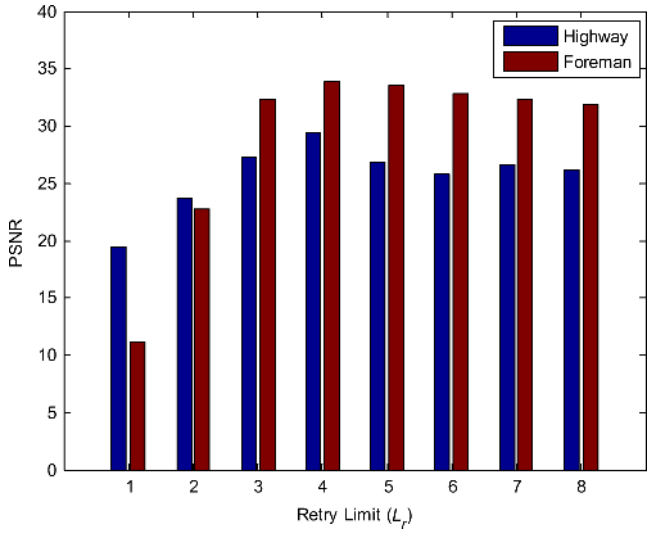

Fig. 11. PSNR results for different values of retry limit from real video NS-2 simulation.

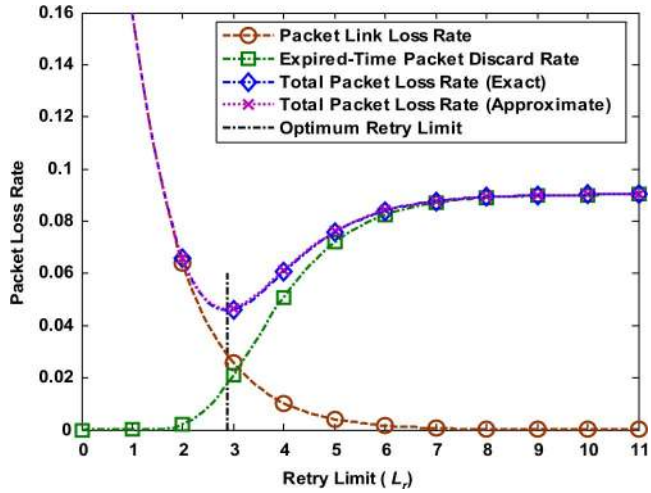

Fig. 12. "Packet link loss rate", the "expired-time packet discard rate" from (17), and the "total packet loss rate" from (37) and (38) versus MAC retry limit along with the value of the L_{rOpt} from (42), in the M/M/1 queuing analysis model ($\lambda = 260$ PPS, $\mu_0 = 453$ PPS, $P_e = 0.4$, and $T_{ex} = 0.2$ s).

expression for the optimum retry limit and design a simple algorithm for retry-limit adaptation in the next section.

V. Optimum Retry Limit and Theoretical Discussions

The analysis and simulation results in Figs. 9 and 10 showed that for an optimum L_r value, throughput is maximized and the "total packet loss rate", p_T, is minimized. We also showed that by setting the buffer size to its optimum value, K_{Opt}, we can avoid "packet overflow drop" in the system. Here we aim to find the optimum retry limit in order to minimize the "total packet loss rate". Hence, we will first solve the optimization problem by employing the M/M/1-based model and will derive a closed-form expression for optimum retry-limit.

A. Optimum Retry Limit

We showed that with $K_A = K_{AOpt}$, the "total packet loss rate" could be calculated from (36) as

$$p'_{TA} = p_{ex} + p_L - p_{ex} \cdot p_L \tag{37}$$

where p_L and P_{ex} are obtained from (1) and (17), respectively.

Assuming that p_L and p_{ex} are relatively small such that the term $p_{ex} \cdot p_L$ is negligible, the "total packet loss rate" (P'_{TA}) can be approximated as

$$p'_{TA} \cong p_{ex} + p_L. \tag{38}$$

The "packet link loss rate" in (1), the "expired-time packet discard rate" in (17), and the "total packet loss rate" in (37) and (38) are plotted in Fig. 12. It can be seen that the difference between the two latter curves is negligible. Therefore, for simplicity, we use the approximate expression (38) for the "total packet loss rate" in order to derive the optimum retry-limit. The "total packet loss rate" in M/M/1 queuing model, $p'_{TA_{(M)}}$, under the above assumptions can be obtained by using (1) and (17) in (38) as

$$\begin{aligned} p'_{TA_{(M)}} &= p_{ex_{(M)}} + p_L \\ &= \left(\rho_0 \frac{1 - P_e^{L_r+1}}{1 - P_e}\right) e^{-\left(\mu_0(1-P_e)/1-P_e^{L_r+1}-\lambda\right)T_{ex}} \\ &\quad + P_e^{L_r+1}. \end{aligned} \tag{39}$$

Now by solving $dp'_{TA_{(M)}}/dL_r = 0$, the optimal L_r ($L_{rOpt_{(M)}}$) in our M/M/1 model can be obtained from the following equation:

$$L_{rOpt_{(M)}} = -1 + \log_{P_e}\left(1 - \frac{T_{ex}\mu_0(1-P_e)}{T_{ex}\lambda + Ln(\beta)}\right) \tag{40}$$

where β is computed as

$$\beta = \frac{\rho_0}{1 - P_e} + \frac{T_{ex}\lambda}{1 - P_e^{L_{rOpt_{(M)}}+1}}. \tag{41}$$

In the recursive formula (40), the value of $P_e^{L_r+1}$ (in β) will be assumed to be negligible compared to one. Then the optimum L_r can be approximately written as

$$L_{rOpt_{(M)}} \cong -1 + \log_{P_e}\left(1 - \frac{T_{ex}\mu_0(1-P_e)}{T_{ex}\lambda + Ln(\frac{\rho_0}{1-P_e} + T_{ex}\lambda)}\right). \tag{42}$$

The value of $L_{rOpt_{(M)}}$ from (42) is shown as the vertical line in Fig. 12. This figure shows the accuracy of the approximate optimum retry limit, $L_{rOpt_{(M)}}$, which is obtained from (42).

B. Considerations on the Analysis of Retry-Limit Optimization

In the following, we derive the upper and lower bounds for P_e according to our M/M/1-based analysis. In order to obtain

an acceptable and real-valued answer to $dp_{T_{(M)}}/dL_r = 0$, the following conditions must be met:

Constraint 1: The first constraint is a direct result of inequality $\rho = \lambda/\mu < 1$. Following our arguments in [22] in deriving the PER bounds, we have

$$P_{eH} < 1 - \rho_0 \tag{43}$$

where $\rho_0 = \lambda/\mu_0$ and P_{eH} is the upper-bound of P_e. The above approximation is good for retry limit values as low as $L_r = 8$. Given that $L_r = 8$ (and with $\lambda = 260$ PPS and $\mu_0 = 453.6$ PPS), the corresponding value of P_{eH} from (43) will be 0.4299.

Constraint 2: In order to obtain a positive real value in (42), we need to satisfy

$$\left(1 - \frac{T_{ex}\mu_0(1-P_e)}{T_{ex}\lambda + Ln(\frac{\rho_0}{1-P_e} + T_{ex}\lambda)}\right) > 0. \tag{44}$$

In (44), it can be shown numerically that, for practical conditions, the value of $\rho_0/(1-P_e)$ is negligible compared to $T_{ex}\lambda$ (e.g., for $\mu_0 = 453.6$ PPS, $T_{ex} = 0.2$ PPS, $\lambda = 260$ PPS, and $P_e = 0.4$, we have $\rho_0/(1-P_e) = 0.95$ and $T_{ex}\lambda = 52$); hence, we have the following lower-bound, P_{eL}, for P_e:

$$P_{eL} > 1 - \frac{\lambda}{\mu_0} - \frac{Ln(T_{ex}\lambda)}{T_{ex}\mu_0}. \tag{45}$$

The lower bound for $P_e(P_{eL})$ (e.g., with $\lambda = 260$ and $P_e = 0.4$) corresponds to the case where the minimum packet loss occurs at infinity. The two constraints, (43) and (45), can be combined for M/M/1-based model as

$$\left(1 - \frac{\lambda}{\mu_0} - \frac{Ln(T_{ex}\lambda)}{T_{ex}\mu_0}\right) < P_e < (1-\rho_0). \tag{46}$$

Setting $\mu_0 = 453.6$ PPS and $T_{ex} = 0.2$ s in (46), P_e will be bounded within $0.3832 < P_e < 0.4268$ and $0.2934 < P_e < 0.3386$, respectively, for $\lambda = 260$ PPS and $\lambda = 300$ PPS.

C. Delay: Simulation and Analysis

For analyzing the packet delay performance as a function of retry limit, we have implemented a scenario using NS2 where video packets are sent at the average rate of $\lambda = 152$ PPS thorough a wireless network. In this scenario, buffer length and expiry time are set to $K_A = 50$ and $T_{ex} = 0.2$, respectively. In this real-time simulation, the sending and receiving times for each packet are recorded. The simulations are run for each value of L_r and the corresponding average delay times are calculated and shown in Fig. 13. Average delay time here is defined as the time difference between sending time by the wireless access point and receiving time at the wireless node. As seen from Fig. 13, packet delay increases with retry limit up to a certain point, which is about $T_1 = 0.125$ s. For a queuing system with packet expiry as the only source of packet loss, we would expect this certain amount of delay to be close to expiry time $T_{ex} = 0.2$. However, since buffer overflow drop is another cause of packet loss in our simulations, there is a total effective buffer size due to combined effect of the physical and virtual buffers.

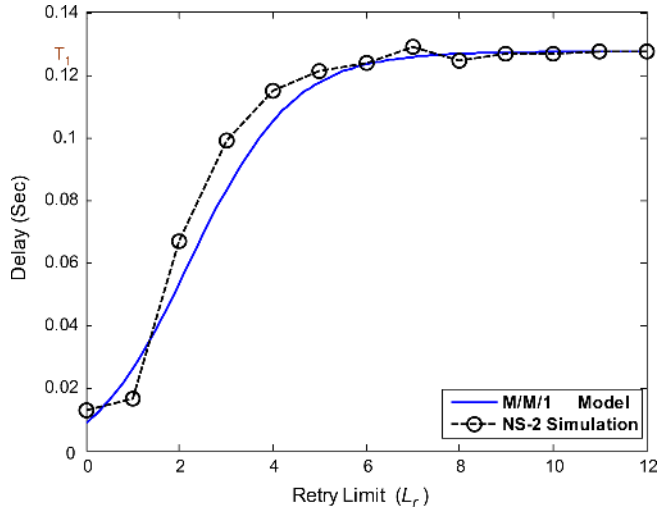

Fig. 13. Delay versus MAC retry limit, based on NS2 simulation and M/M/1 Model ($\lambda = 152$ PPS, $P_e = 0.4$, $K_A = 50$, and $T_{ex} = 0.20$ s).

To obtain this effective buffer size, which is shorter than both physical and virtual buffers, we substitute T_1 in (30), i.e., we set $T_q = T_1$. The corresponding effective buffer size will be $K_{A_{Total}} = 19$. Also the physical buffer size is set to $K_A = 50$ and from (32), the virtual buffer size is derived $K'_A = 30.4$. Our numerical results show that the general relation between real buffer length, virtual buffer length, and total effective buffer length can be shown as

$$K_{A_{Total}} = \frac{K_A K'_A}{K_A + K'}. \tag{47}$$

In Fig. 13, we have also plotted packet delay based on our M/M/1 queuing model, which quite closely matches our simulation-based result. In deriving this curve, we have used the following formula [23, Section 2.2.4]:

$$W = W_q + \frac{1}{\mu} = \frac{\rho}{\mu - \lambda} + \frac{1}{\mu}. \tag{48}$$

VI. Algorithm

In this section, we propose a new algorithm for optimizing the number of retransmissions based on the M/M/1 analytical model. This algorithm will provide a practical means to implement our adaptive retry-limit method in real-time systems. In order to clarify the logic behind the algorithm, we rewrite the solution to $dp_{T_{(M)}}/dL_r = 0$ as follows where $P_{T_{(M)}}$ is from (39):

$$\rho e^{-\left(\frac{\mu_0(1-P_e)}{1-P_e^{L_r+1}} - \lambda\right)T_{ex}} = \rho\left(\frac{\rho_0}{1-P_e} + \frac{T_{ex}\lambda}{1-P_e^{L_r+1}}\right)^{-1}. \tag{49}$$

The LHS of (49) is equal to the "expired-time packet discard rate" at $L_r = L_{rOpt}$, which we denote by p_{exOpt}. Considering $\rho = \lambda/\mu$, we can write

$$p_{exOpt} = \frac{1 - P_e^{L_r+1}}{1 + \frac{T_{ex}\mu_0(1-P_e)}{1-P_e^{L_r+1}}}. \tag{50}$$

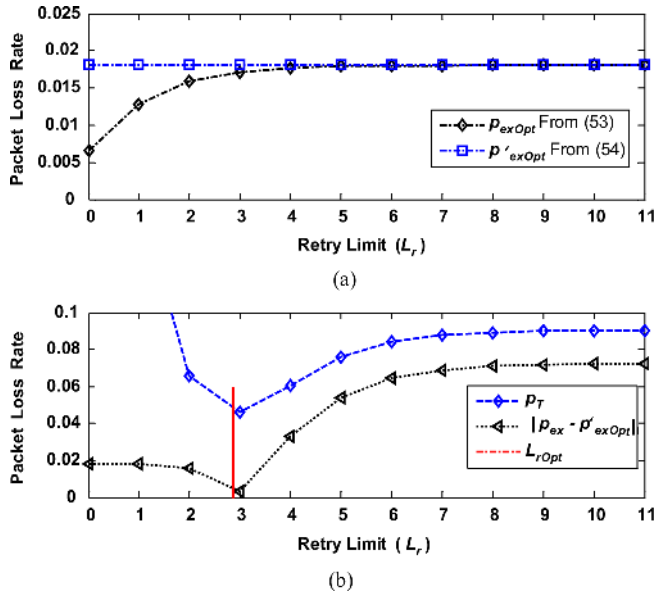

Fig. 14. P_{ex}, P_{exOpt}, p'_{exOpt}, P_T, and L_{rOpt} along with $|P_{ex} - P_{exOpt}|$ versus L_r, in M/M/1 model ($\lambda = 260$ PPS, $\mu_0 = 453$ PPS, $P_e = 0.4$, and $T_{ex} = 0.2$ s).

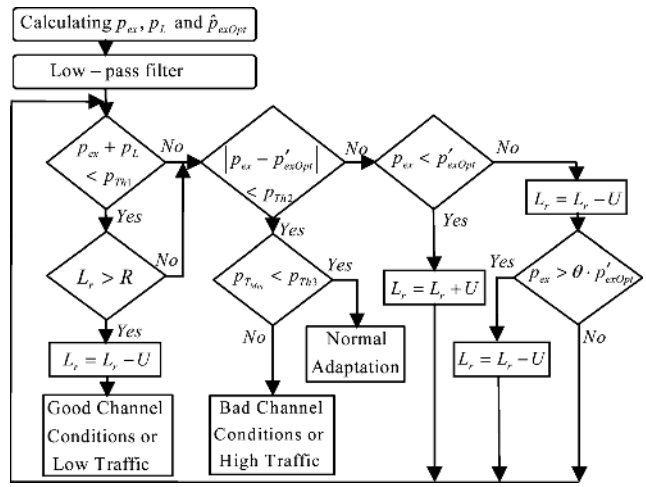

Fig. 15. Adaptive algorithm flowchart ($p_{Th1} = 0.02$, $p_{Th2} = 0.005$, $p_{Th3} = 0.1$,R $= 5$, $\theta = 10$, U $= 1$).

Assuming that $P_e^{L_r+1}$ is negligible compared to one, (50) can be approximately written as

$$p'_{exOpt} = \frac{1}{1 + T_{ex}\mu_0(1 - P_e)}. \tag{51}$$

The values of p_{exOpt} and p'_{exOpt}, respectively, from (50) and (51) versus retry-limit (L_r) are shown in Fig. 14(a). Also, the values of $p_{T_{(M)}}$ and L_{rOpt}, respectively, from (39) and (42) and the value of $|p_{ex_{(M)}} - p'_{exOpt}|$ are shown in Fig. 14(b) versus retry-limit (L_r). These figures clearly show that p'_{exOpt} is a fairly accurate approximation for our purposes. According to Fig. 14(b), we have

$$p_{ex_{(M)}} - p'_{exOpt} \begin{cases} < 0 & L_r < L_{ropt} \\ = 0 & L_r = L_{ropt} \\ > 0 & L_r > L_{ropt}. \end{cases} \tag{52}$$

Based upon the above observations, the flowchart for the algorithm is given in Fig. 15. The algorithm requires checking the amount of minimum "total packet loss rate" "$p_{T_{Min}}$" to ensure that p_T is within the desired range. Therefore, we assume that the acceptable value of $p_{T_{Min}}$ is less than p_{Th3}.

The threshold value p_{Th1} is employed in the algorithm flowchart to fulfill the second constraint of Section V-B.

Actually condition $p_{ex} + p_L < p_{Th1}$ corresponds to good channel or light input traffic. When this condition is fulfilled, the algorithm proceeds by decreasing L_r to settle below a certain value R. By setting $P_e = P_{eL}$ and $L_r = R$, the value of p_{Th1} can be calculated from (39) as follows:

$$p_{Th1} = \left(\rho_0 \frac{1 - P_{eL}^{R+1}}{1 - P_{eL}}\right) e^{-\left(\frac{\mu_0(1-P_{eL})}{1-P_{eL}^{R+1}} - \lambda\right)T_{ex}} + P_{eL}^{R+1}. \tag{53}$$

For $R >> 1$, (53) can be simplified to

$$p_{Th1} \cong \frac{\rho_0 e^{-\mu_0(1-P_{eL}-\rho_0)T_{ex}}}{1 - P_{eL}}. \tag{54}$$

In the numerical results, we have set R $= 5$. With this value of R, and for $P_{eL} = 0.3832$, $\lambda = 260$ PPS, $\mu_0 = 453.6$ PPS, and $T_{ex} = 0.2$, p_{Th1} will be obtained from (53) and (54) equal to 0.02 and 0.0178, respectively. In order to satisfy constraint 2 with a margin, we have set $p_{Th1} = 0.02$, which is slightly larger than the value of p_{Th1} derived from (53). We have assessed the performance of this algorithm through simulations. P_e is assumed to be a normally-distributed random variable and the optimum L_r is obtained via exploiting the algorithm of Fig. 15 for each sample of P_e, which corresponds to a given channel state. Although the assumption in simulating the algorithm is "Normal distribution" for PER, our algorithm can be used for any other distribution as well, since it deals with estimated PER regardless of the statistical properties of the PER variations. The simulation results corresponding to a two-state PER distribution was also shown in Section IV-E.

Numerical analysis shows that setting $p_{Th1} = 0.02$, $p_{Th2} = 0.005$, $p_{Th3} = 0.1$, and R $= 5$ can provide the optimal set of control parameters in the algorithm.

The throughput for a system with no L_r adaptation is provided in Fig. 16(a) along with the achievable throughout from the above L_r adaptation algorithm. This figure shows that adaptation of retry-limit in response to the channel condition can remarkably improve the system throughput. According to Fig. 16(a), the adaptive algorithm for this example has increased the average throughput from 140 PPS to 230 PPS. The corresponding packet loss rate and packet delay performance of our adaptive real-time algorithm are also shown, respectively, in Fig. 16(b) and (c).

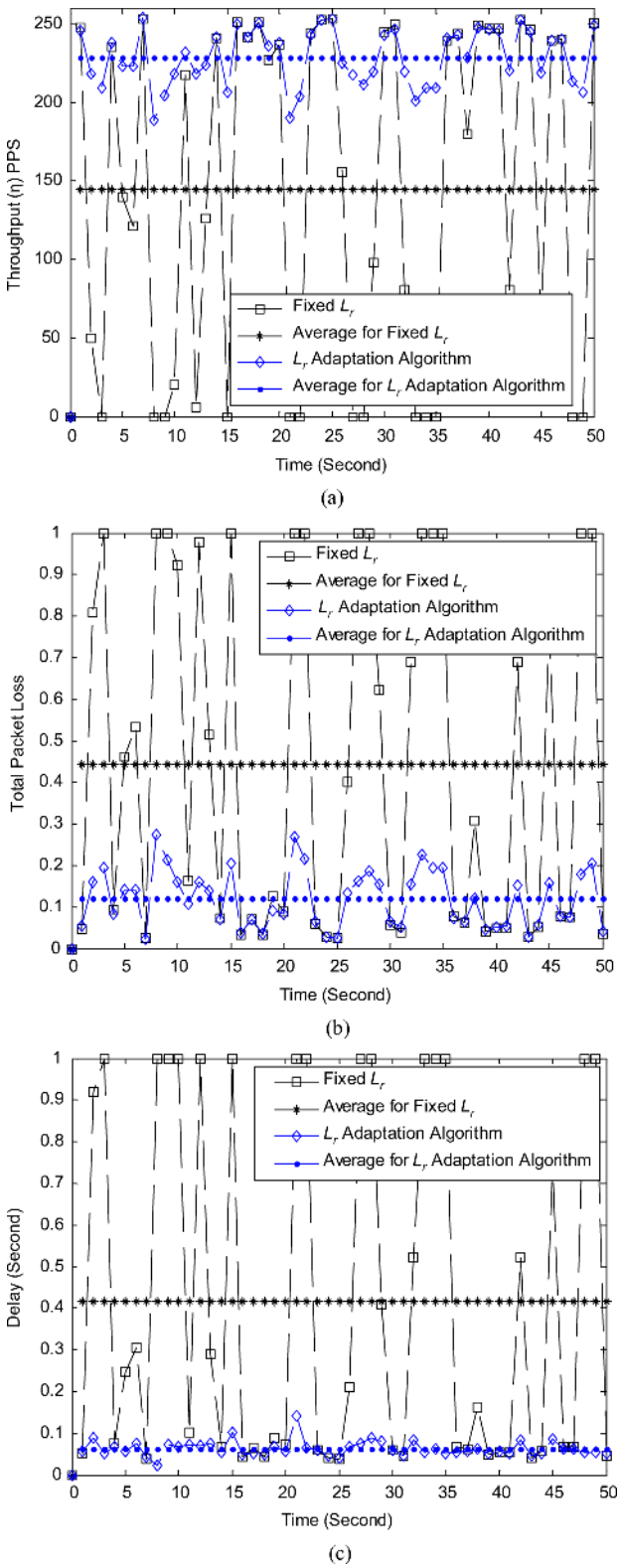

Fig. 16. Comparing results for the fixed retry-limit and adaptive algorithm. (a) Throughput. (b) Total packet loss. (c) Delay.

VII. Conclusion

We have analyzed a QoS provisioning strategy for delay-sensitive and real-time applications such as live video streaming over WLAN. New mathematical expressions are derived for "expired-time packet discard rate" by employing M/G/1, M/M/1/K, and M/M/1 models in a cross-layer platform. First, these models are studied in the presence of two sources of packet loss: "expired-time packet discard" and "packet link loss" mechanisms. We have particularly shown that the M/M/1 model is a simple and accurate model for analyzing delay-limited applications such as live video transmission over WLAN. Then we considered the case where "packet overflow drop" is also present in the system as a third source of packet loss. Subsequently, by showing the interaction between "packet overflow drop" and "expired-time packet discard" mechanisms in the queue and by introducing the "virtual buffer size" concept, we showed that the expired-time mechanism could be modeled by a secondary "packet overflow drop" process. It was deduced that "packet overflow drop" in the queue can be prevented by appropriate setting of the buffer size. Using both mathematical analysis and NS-2 simulations with exponential and real video traffic, we derived the optimal MAC retry limit value for which the "total packet loss rate" is minimized. Finally, we used M/M/1 queuing to introduce a new adaptive retry-limit algorithm for enhancing delay-sensitive video quality over IEEE 802.11 WLANs.

Appendices

Appendix A:

Deriving $W_{(G)}$: Substituting B^* from (14) into (13) and using algebraic manipulations, (13) can be written as

$$W^*(s) = \frac{\sum_{i=0}^{L_r} b_i s^{L_r - i}}{\sum_{i=0}^{L_r} a_i s^{L_r - i}} \tag{A1}$$

where b_i and a_i coefficients can be obtained by expanding (A1) for each value of the L_r. In Appendix B, we have obtained the coefficients b_i and a_i for $L_r = 1, 2, 3$, and 4. Then, we can rewrite $W^*(s)$ in the form of the sum of partial fractions as

$$W^*(s) = \kappa + \sum_{i=0}^{L_r} \frac{r_i}{s - \alpha_i} \tag{A2}$$

where $\kappa = b_0/a_0, \alpha_0 \cdots \alpha_{L_r}$, are the denominator's root in (A1) and r_1 can be calculated as

$$r_i = (s - \alpha_i) W_q^*(s)|_{s=\alpha_i}. \tag{A3}$$

Now from inverse Laplace transform of (A2), we derive the service time PDF as

$$w(x) = \kappa\delta(x) + \sum_{i=0}^{L_r} r_i(e^{-\alpha_i x}). \tag{A4}$$

Service time PDF can be obtained from the integral of $W(x)$ in (A4):

$$W(x) = \kappa u(x) + \sum_{i=0}^{L_r} \frac{r_i}{\alpha_i}(1 - e^{-\alpha_i x}). \tag{A5}$$

In (A5) for $T_{ex} > 0$, we have $\kappa u(x) = \kappa$; hence

$$W(T_{ex}) = \kappa + \sum_{i=0}^{L_r} \frac{r_i}{\alpha_i}(1 - e^{-\alpha_i T_{ex}}). \tag{A6}$$

APPENDIX B:

$$L_r = 0 \quad W_q^*(s) = \frac{(1-\rho)(s+\mu_0)}{s+\mu_0-\lambda} \tag{B1}$$

$$L_r = 1 \quad W_q^*(s) = \frac{2(1-\rho)(s+\mu_0)(s+\frac{\mu_0}{2})}{2s^2+(3\mu_0-2\lambda)s-\mu_0\lambda(1+P_e)+\mu_0^2} \tag{B2}$$

$$L_r = 2 \quad W_q^*(s) = \frac{6(1-\rho)(s+\mu_0)(s+\frac{\mu_0}{2})(s+\frac{\mu_0}{3})}{\begin{pmatrix} 6s^3+(11\mu_0-6\lambda)s^2 \\ -\left[\mu_0\lambda(5+3P_e+P_e^2)-6\mu_0^2\right]s \\ -\mu_0^2\lambda(1+P_e+P_e^2)+\mu_0^3 \end{pmatrix}}. \tag{B3}$$

REFERENCES

[1] IEEE Standard 802.11-1999, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1999.

[2] A. Shojaeifard, F. Zarringhalam, and M. Shikh-Bahaei, "Joint physical layer and data link layer optimization of CDMA-based networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3278–3287, Oct. 2011.

[3] A. Shadmand, K. Nehra, and M. Shikh-Bahaei, "Cross-layer design in dynamic spectrum sharing systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, Article ID 458472, 2010.

[4] L. Ong, M. Shikh-Bahaei, and J. A. Chambers, "Variable rate and variable power MQAM system based on Bayesian bit error rate and channel estimation techniques," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 177–182, Feb. 2008.

[5] H. Bobarshad and M. Shikh-Bahaei, "M/M/1 queuing model for adaptive cross-layer error protection in WLANs," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Apr. 2009.

[6] M. Van der Schaar and N. S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms," *Wireless Commun.*, vol. 12, pp. 50–58, Aug. 2005.

[7] S. Hua, Y. Guo, Y. Liu, H. Liu, and S. S. Panwar, "Scalable video multicast in hybrid 3G/AD-Hoc networks," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 402–413, Apr. 2011.

[8] S. Han, H. Joo, D. Lee, and H. Song, "An end-to-End virtual path construction system for stable live video streaming over heterogeneous wireless networks," *IEEE J. Select. Areas Commun.*, vol. 29, no. 5, pp. 1032–1041, May 2011.

[9] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE," *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 102–109, Feb. 2010.

[10] M. van der Schaar and D. S. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 185–197, Jan. 2007.

[11] N. S. Shankar and M. van der Schaar, "Performance analysis of video transmission over IEEE 802.11a/e WLANs," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2346–2362, Jul. 2007.

[12] M. van der Schaar, S. Krishnamachari, C. Sunghyun, and X. Xiaofeng, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE J. Select. Areas Commun.*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.

[13] M. van der Schaar, Y. Andreopoulos, and H. Zhiping, "Optimized scalable video streaming over IEEE 802.11 a/e HCCA wireless networks under delay constraints," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 755–768, Jun. 2006.

[14] M. van der Schaar, D. S. Turaga, and W. Raymond, "Classification-based system for cross-layer optimized wireless video transmission," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1082–1095, Oct. 2006.

[15] M. Lu, P. Steenkiste, and T. Chen, "Video streaming over 802.11 WLANs with content-aware adaptive retry," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, Jul. 2005.

[16] M. M. Krunz and J. G. Kim, "Fluid analysis of delay and packet discard performance for QoS support in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 19, no. 2, pp. 384–395, Feb. 2001.

[17] H. Lee, S. Lee, and A. C. Bovik, "Cross-layer optimization for downlink wavelet video transmission," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 813–823, Aug. 2011.

[18] P. U. Tournoux, E. Lochin, J. Lacan, A. Bouabdallah, and V. Roca, "On-the-Fly erasure coding for real-time video applications," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 797–812, Aug. 2011.

[19] Q. Li and M. van der 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.

[20] T. H. Luan, L. X. Cai, and X. Shen, "Impact of network dynamics on user's video quality: Analytical framework and QoS provision," *IEEE Trans. Multimedia*, vol. 12, no. 1, pp. 67–78, Jan. 2011.

[21] J.-L. Hsu and M. van der Schaar, "Cross layer design and analysis of multiuser wireless video streaming over 802.11e EDCA," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 268–271, Apr. 2009.

[22] H. Bobarshad, M. van der Schaar, and M. R. Shikh-Bahaei, "A low-complexity analytical modelling for cross-layer adaptive error protection in video over WLAN," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 427–438, Aug. 2010.

[23] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. New York: Wiley, 1998.

[24] Q. Li, Y. Andreopoulos, and M. van der Schaar, "Streaming-viability analysis and packet scheduling for video over QoS-enabled networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3533–3549, Nov. 2007.

[25] The network Simulator 2 (NS-2), Information Sciences Institute (ISI), Roadmap—User Information—Developer Information—Projects—Contributing. [Online]. Available: http://nsnam.isi.edu/nsnam/index.php/User_Information.

[26] L. Kleinrock, *Queueing Systems Volume 1: Theory' 1975*. New York: Wiley-Interscience, 1975.

Hossein Bobarshad (M'10) received the M.Sc. degree in communication engineering from the Iran University of Science & Technology (IUST), Tehran, Iran, and the Ph.D. degree in wireless networking from the Electronic Engineering Department of King's College London, London, U.K., in 2010.

He is currently a Lecturer at the University of Tarbiat Modares, Tehran, Iran. He has worked as research assistant at King's College London. His main research interests are in the areas of data communication theory, mathematical modeling, and real-time video transmission through unreliable media, e.g., wireless networks and the Internet. In particular, he conducts his research mainly on cross-layer design for adaptive error protection in video over WLAN and wireless sensor networks for monitoring of physiological signals.

Mihaela van der Schaar (F'10) is currently a Professor in the Electrical Engineering Department at University of California, Los Angeles. She holds 33 granted U.S. patents and three ISO awards for her contributions to the MPEG video compression and streaming international standardization activities. Her research interests are in multimedia communications, networking, processing, and systems and, more recently, on learning and games in engineering systems.

Prof. van der Schaar received in 2004 the NSF Career Award, in 2005 the Best Paper Award from the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, in 2006 the Okawa Foundation Award, in 2005, 2007 and 2008 the IBM Faculty Award, and in 2006 the Most Cited Paper Award from *EURASIP: Image Communications* journal. She is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON MULTIMEDIA. She was an associate editor for the IEEE TRANSACTIONS ON MULTIMEDIA, IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SIGNAL PROCESSING MAGAZINE, etc.

A. Hamid Aghvami (F'05) is presently the Director of the Centre for Telecommunications Research at King's College London, London, U.K. He joined the academic staff at King's in 1984. In 1989, he was promoted to Reader, and in 1993 was promoted to Professor in Telecommunications Engineering. He has published over 500 technical papers and given invited talks and courses the world over on various aspects of personal and mobile radio communications. He is founder of the International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), a major yearly conference attracting some 1000 attendees.

Prof. Aghvami is a Fellow of the Royal Academy of Engineering, Fellow of the IET, and in 2009 was awarded a Fellowship of the Wireless World Research Forum in recognition of his personal contributions to the wireless world, and for his research achievements as Director at the Centre for Telecommunications Research at King's College London.

Reza S. Dilmaghani (M'06) received the Ph.D. degree from King's College London, London, U.K., in June 2006.

Since May 2006, he has been a Lecturer with the Ultra wideband Communications Group. His interests include signal processing for communications, (medical) image processing, and wavelet theory. He has also published several technical papers in recognized conferences and journals.

Dr. Dilmaghani was awarded the esteemed European Information Society Technologies (IST) prize sponsored by the European Commission in 2005 which is the most prestigious technology prize for innovative products all over Europe.

Mohammad R. Shikh-Bahaei (SM'08) received the M.Sc. degree in electronic engineering from Sharif University of Technology, Tehran, Iran, and the Ph.D. degree in wireless communications from King's College London, London, U.K., in 2000.

He has worked for Algorex Ltd., U.K. (now part of National Semiconductor Corp., CA, USA) and for a start-up company, MobilSoft Ltd., U.K. In December 2000, he joined National Semiconductor Corp. (NSC), CA, USA, and worked on the design of 3rd-generation handsets based on UMTS standards, for which he has been awarded three U.S. patents as inventor and co-inventor, respectively. He is currently a Senior Lecturer within the Centre for Telecommunications Research at the Department of Informatics at King's College London. The main focus of his research has been on the analysis and design of adaptive transmitters and resource allocation techniques in wireless communications, and optimization of wireless network performance for services such as VoIP and Video-over-IP. He was the Lead Editor of a special Issue of *EURASIP Journal on Wireless Communications and Networking*. He is also the founder and organizer of the annual international IEEE conference Wireless Advanced (formerly known as SPWC) which is supported by CTR of King's College London.