Wireless Netw
DOI 10.1007/s11276-008-0137-y

An analytical model for performance evaluation of multimedia applications over EDCA in an IEEE 802.11e WLAN

Sri Harsha · Anurag Kumar · Vinod Sharma

Abstract We extend the modeling heuristic of (Harsha et al. 2006. In *IEEE IWQoS '06*, pp 178 – 187) to evaluate the performance of an IEEE 802.11e infrastructure network carrying packet telephone calls, streaming video sessions and TCP controlled file downloads, using Enhanced Distributed Channel Access (EDCA). We identify the time boundaries of activities on the channel (called channel slot boundaries) and derive a Markov Renewal Process of the contending nodes on these epochs. This is achieved by the use of attempt probabilities of the contending nodes as those obtained from the saturation fixed point analysis of (Ramaiyan et al. 2005. In *Proceedings ACM Sigmetrics, '05*. Journal version accepted for publication in IEEE TON). Regenerative analysis on this MRP yields the desired steady state performance measures. We then use the MRP model to develop an effective bandwidth approach for obtaining a bound on the size of the buffer required at the video queue of the AP, such that the streaming video packet loss probability is kept to less than 1%. The results obtained match well with simulations using the network simulator, *ns*-2. We find that, with the default IEEE 802.11e EDCA parameters for access categories AC 1, AC 2 and AC 3, the voice call capacity decreases if even one streaming video session and one TCP file download are initiated by some wireless station. Subsequently, reducing the voice calls increases the video downlink stream throughput by 0.38 Mbps and file download capacity by 0.14 Mbps, for every voice call (for the 11 Mbps PHY). We find that a buffer size of 75KB is sufficient to ensure that the video packet loss probability at the QAP is within 1%.

Keywords VoIP on WLAN · Streaming video on WLAN · TCP throughput on WLAN · Capacity of IEEE 802.11e WLAN · Performance modeling of EDCA · Buffer sizing at access point

This is an extended version of our paper (Harsha et al. 2006. An analytical model for the capacity estimation of combined VoIP and TCP file transfers over EDCA in an IEEE 802.11e WLAN, pp. 178–187, 19–21 June 2006) in IEEE IWQoS '06.

S. Harsha · A. Kumar (✉) · V. Sharma
Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore 560012, India
e-mail: anurag@ece.iisc.ernet.in

S. Harsha
e-mail: harshas@ece.iisc.ernet.in

V. Sharma
e-mail: vinod@ece.iisc.ernet.in

1 Introduction

The IEEE 802.11e standard [1] provides service differentiation in IEEE 802.11 WLANs, with the introduction of a single coordination function called hybrid coordination function (HCF). HCF combines the distributed coordination function (DCF) and point coordination function (PCF) of IEEE 802.11 MAC for QoS data transmission. In IEEE 802.11e, a superframe still consists of the two phases of operations, contention period (CP) and contention free period (CFP). Enhanced distributed coordination access (EDCA) is used only in the CP, while HCF controlled channel access (HCCA) can be used in both phases. A QoS enabled access point (AP) is called a QAP, whereas a QoS enabled station (STA) is called a QSTA. The HCCA is deterministic and hence yields to simple calculations for performance analysis. The EDCA is based on random access and hence demands stochastic modeling approach.

EDCA offers the possibility of defining four different classes of service at the MAC layer so that QoS requirements of multimedia traffic can be supported in addition to data traffic. At the MAC layer, each service class is called an *access category* (AC), and service between classes is differentiated by different sets of channel contention parameters. See Table 1 for parameters of different ACs. It is through these ACs that the differentiation is achieved.

Performance analysis of IEEE 802.11e WLANs has become an active research area. While many simulation studies have been reported [2–5], it is important to develop analytical models. Analytical modeling provides insights into the working of the system and leads to a more general understanding of the effects of various parameters, and design choices, than many simulation runs. Further, these models may provide general guidelines for admission control and MAC parameter optimization, and may lead to ideas for novel adaptive MAC algorithms. The availability of good analytical models is also useful for developing fast simulations [6–8].

Related Literature: Model based performance analysis of EDCA 802.11e WLANs have been proposed in [9–14]. Robinson and Randhawa [11], Zhu and Chlamtac [12] and Kong et al. [13] consider a WLAN with saturated nodes (nodes that always have packets to transmit). Ramaiyan et al. [9] extend the fixed point analysis of Kumar et al. [15] for a single cell IEEE 802.11e WLAN with saturated nodes and propose a general fixed point analysis that captures the differentiation by minimum contention window (CW), maximum CW and arbitration interframe space (AIFS).

With traffic from actual applications, however, the nodes are not always saturated. Shankar et al. [14] evaluate the VoIP capacity in 802.11e WLAN, but in a scenario where other classes of traffic are not coexistent in the WLAN. Clifford et al. [16] have proposed a model for 802.11e for different classes of traffic when the nodes are non saturated. This model yields throughputs of various flows. The authors do not model the buffer dynamics for different traffic types.

Our Contribution: We extend our heuristic model in [17] to predict the performance of a single cell infrastructure IEEE 802.11e WLAN, under a scenario where VoIP traffic, downlink streaming video sessions and TCP controlled data download traffic are carried over EDCA. Then, by applying the effective bandwidth approach, we use the derived model to obtain design insights of the size of buffer required for the AC 2 queue at the QAP. In both the cases, the analytical results closely match with the simulation results. We establish the fact that the heuristic of using saturation attempt probabilities in a non saturated scenario is an effective approach and can be applied widely to obtain various performance metrics of the system.

Table 1 Parameters of different ACs as defined in 802.11e

Access category	CW_{min}	CW_{max}	AIFS	TXOP[a] max limit	Usage
AC(3)	7	15	2	3.264 ms	Voice
AC(2)	15	31	2	6.016 ms	Video
AC(1)	31	1,023	3	–	Best effort
AC(0)	31	1,023	7	–	Background

[a] For 802.11b PHY

Paper Outline: In Sect. 2 we discuss the approach for modeling along with the observations and assumptions of the network and the traffic. In Sect. 3 we formulate a Markov renewal framework, by using the state dependent attempt probabilities of [9]. In Sect. 4 we derive the performance measures, namely, the VoIP call capacity, saturation video throughput and the aggregate TCP throughput. In Sect. 5, we present further analysis of streaming video sessions and obtain the service time distribution of video packet successes. By an 'effective bandwidth approach', we find the video buffer size required at the access point (AP), to meet the packet loss QoS. In Sect. 6 we present the numerical and simulation results for all the measures so derived. Lastly, in Sect. 7 we conclude with the listing of useful modeling and performance insights obtained in this analysis.

2 The modeling approach

We study the performance of a single cell infrastructure 802.11e WLAN that uses EDCA, when AC 3, AC 2 and AC 1 are used for voice, video and data respectively. The modeling approach follows that of [17] and can be briefly explained as follows:

(1) Embed the number of contending nodes (i.e., those that have non empty queues) at *channel slot boundaries*. The *channel slot boundaries* are those instants of time when an activity ends or there is a back off slot after which no node attempts. The activity could be a successful transmission or a collision.

(2) Use the heuristic that, if n nodes are contending at a channel slot boundary, their attempt probabilities are those obtained from fixed point analysis of [9] with n saturated nodes.

(3) Use the thus obtained attempt probabilities to model the evolution of the number of contending nodes at channel slot boundaries. Since the channel slot durations depend on the activity, this yields a Markov renewal process [18, Chapter 2].

(4) Obtain the stationary probability vector π of the embedded Markov chain of the Markov renewal process.

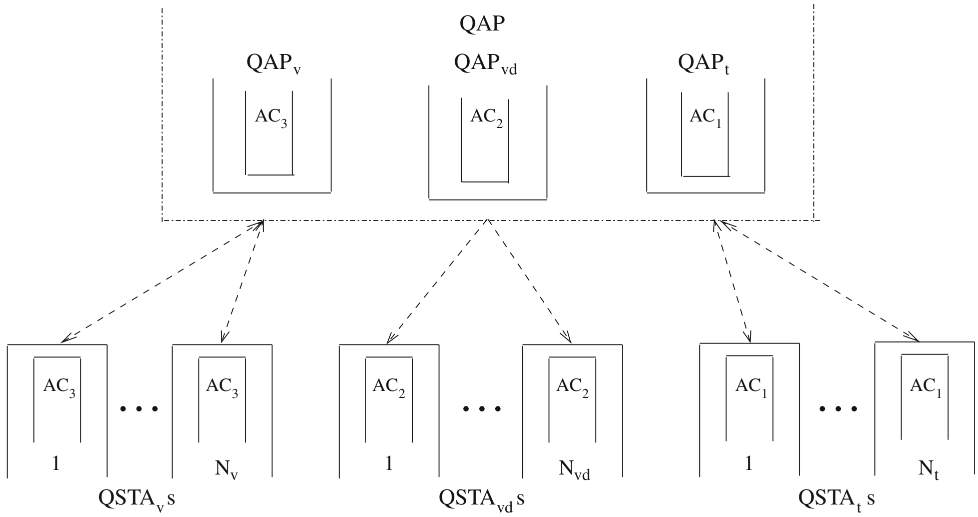

Fig. 1 An IEEE 802.11e WLAN model scenario where VoIP calls, streaming video sessions and TCP traffic are serviced on EDCA

(5) Use a Markov regenerative argument to obtain the performance measures [18, 19].

2.1 The network scenario and modeling observations

We consider an infrastructure IEEE 802.11e WLAN, which has VoIP, downlink video streaming and TCP controlled file download traffic, serviced on EDCA. While IEEE 802.11e also defines EDCA TXOPs for transmission of more than one MSDUs (MAC Service Data Unit) when a node obtains the opportunity to transmit [1, Section 9.1.3.1], we use the default value that the sender can send not more than one MSDU in an EDCA TXOP. Let N_v be the number of full duplex CBR VoIP calls, N_{vd} be the number of simplex CBR download video streaming sessions and N_t be the number of TCP controlled file transfers in the WLAN. We carry forward the following assumptions from [17]:

A1 There are no hidden nodes in the WLAN, there are no bit errors, and packets in the channel are lost only due to collisions.

A2 The VoIP traffic, video streaming traffic and TCP traffic all originate from different *QSTA*s. This implies that each *QSTA* has only one type of traffic. Denote the *QSTA*s with VoIP traffic (AC 3 queue) as $QSTA_v$, the *QSTA*s with video streaming traffic (AC 2 queue) as $QSTA_{vd}$ and *QSTA*s with TCP controlled file transfers (AC 1 queue) as $QSTA_t$.

A3 The *QAP* can be viewed as three nodes: QAP_v, an AC 3 queue, for downlink VoIP traffic of all VoIP calls, QAP_{vd}, an AC 2 queue, for downlink video streaming traffic of all video streaming sessions, and QAP_t, an AC 1 queue, for all TCP downloads.

Assumptions A2 and A3 are simplifying implications of an important observation in [9], viz, "with increase in the number of nodes, the performance of the *multiple queues per node* case coincides with the performance of the *single queue per node* case, each node with one queue of the original system". This model is illustrated in Fig. 1. Note that at any time the WLAN in Fig. 1 can be seen to consist of $N_v + N_{vd} + N_t + 3$ nodes.

2.2 VoIP traffic

We consider non-synchronized CBR duplex VoIP calls from codecs that generate VoIP packets every 20 ms. As a QoS requirement we demand that the probability that a packet is transmitted successfully within 20 ms is close to 1 (see [20] for justification). Following are the assumptions that we carry forward from [17] and are justified in [17] and [20]:

A4 The buffer of every $QSTA_v$ has a queue length of at most one packet

A5 New packets arriving to the $QSTA_v$s arrive only at empty queues. This assumption implies that if there are k $QSTA_v$s with voice packets then, a new voice packet arrival comes to a $(k+1)$th $QSTA_v$.

A6 QAP_v is the capacity bottleneck for voice traffic, since, there can be up to N_v packets of different calls in the QAP_v. Therefore to obtain the VoIP capacity of the WLAN, we consider QAP_v saturated. But when we need to evaluate the throughputs of streaming video sessions and TCP download streams, we model the arriving VoIP traffic at QAP_v.

As mentioned earlier, packets arrive every 20 ms in every stream. We use this model in our simulations. However, since our analytical approach is via Markov chains, to model the VoIP traffic, we assume that the probability that a voice call generates a packet in an interval of length l slots is $p_l = 1 - (1-\lambda)^l$, where λ is obtained as follows.

Each system slot is of 20 µs duration (hereafter denoted as δ). Thus in 1,000 system slots there is one arrival. Therefore, for the 802.11b PHY we take $\lambda = 0.001$. This simplification turns out to yield a good approximation.

2.3 TCP controlled file downloads

Each $QSTA_t$ has a single TCP connection to download a large file from a local file server. Hence, the QAP_t delivers TCP data packets towards the $QSTA_t$s, while the $QSTA_t$s return TCP ACKs. We make the following assumptions as in [17] and [20] (see [17] and [20] for justification):

A7 The QAP_t and the $QSTA_t$s have buffers large enough so that TCP data packets or ACKs are not lost due to buffer overflows.

A8 Each $QSTA_t$ can have a maximum of one TCP ACK packet queued up. This assumption implies two things. First, after an $QSTA_t$'s successful transmission, the number of active $QSTA_t$s reduces by one. Second, each successful transmission from the QAP_t activates a new $QSTA_t$.

A9 QAP_t is the traffic bottleneck and hence saturated and always contends for the channel.

2.4 Video streaming traffic

We consider the scenario where the WLAN users connect to a video streaming server located in the wired network, through the QAP.

A10 In our work, we assume that video packets are streamed over UDP between the streaming server and the wireless playout station, without any feedback traffic from the playing station. This assumption implies that the QTA_{vd}s do not have any uplink traffic and hence never contend for the channel.

Li et al. [21] have studied the two dominant streaming multimedia products, RealNetworks *RealPlayer*™ and Microsoft *MediaPlayer*™ and their experiments for a low rate video stream using UDP show that

(1) The sizes of MediaPlayer packets are concentrated around the mean packet size (of 900 bytes). The sizes of RealPlayer packets are spread more widely over a range from 0.6 to 1.8 of the mean normalized packet size.

(2) The packet inter arrival times for RealPlayer varied over a range of 10–160 ms. In contrast, the packet inter arrival times for MediaPlayer are concentrated near 130 ms, indicating that most packets arrive at constant time intervals. The packet inter arrival times were mainly attributed to the property of the streaming server.

Thus they draw the conclusion that the packet sizes and rates generated by MediaPlayer are essentially CBR while the packet sizes and rates generated by RealPlayer are more varied.

A11 In the analysis we obtain the maximum service rate obtainable by the video streams by considering that the video queue is saturated. Thus QAP_{vd} is saturated and always contends for the channel.

A12 In simulations, we consider CBR video streams (one of the two choices as observed by Li et al., discussed above) and consider a rate of 1.5 Mbps and packet size of 1,500 bytes, for validation, since, when the SD-TV (Standard Definition Television) resolution video is coded with H.264 for an MoS (Mean Opinion Score) of 4, the output streaming video rate is 1.5 Mbps (see [22]).

3 The analytical model

3.1 An embedded chain

The evolution of the channel activity in the network is as in Fig. 2. $U_j, j \in 0, 1, 2, 3, \ldots$, are the random instants where either an idle slot, or a successful transmission, or a collision ends. Let us define the time between two such successive instants as a *channel slot*. Thus the interval $[U_{j-1}, U_j)$ is called the jth channel slot. Let the time length of the jth channel slot be L_j (see Fig. 2). The implication of access differentiation through AIFS is that the ACs with larger AIFS values cannot contend in those slots that were preceded by some activity (i.e., successful transmission or collision). After every successful transmission or collision on the channel, AC 1 nodes wait for an additional system slot before contending for the channel. Figure 2 shows the evolution of the channel activity when AC 3, AC 2 and AC 1 queues are active. Note that at the instants U_4, U_6, U_7 and U_{10}, only AC 3 and AC 2 nodes can contend for the channel, whereas AC 1 nodes have still to wait for one more system slot to be able to contend. At other instants, U_5, U_8, U_{11}, U_{12} and U_{13}, all ACs, i.e., AC 3, AC 2 or AC 1 can attempt.

We first consider the case where QAP_v is saturated and contends at all times (see Assumption A6), to obtain the VoIP capacity of the WLAN. Thus QAP_v, QAP_{vd} and QAP_t are always non-empty. We then need to keep track of only non-empty $QSTA_v$s and $QSTA_t$s, to know the number of contending nodes at any channel slot boundary. Let $Y_j^{(v)}$ be the number of non-empty $QSTA_v$s and $Y_j^{(t)}$ be the number of non-empty $QSTA_t$s at the instant U_j. Thus

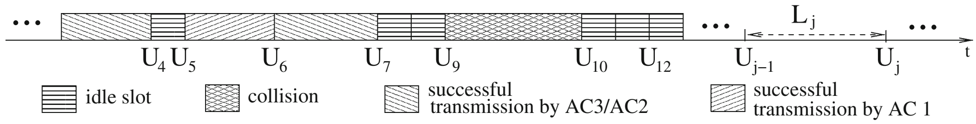

Fig. 2 An evolution of the channel activity with three ACs in 802.11e WLANs. At the instants U_4, U_6, U_7 and U_{10}, only AC 3 and AC 2 can contend for the channel, whereas at other instants, U_5, U_8, U_{11}, U_{12} and U_{13}, all ACs, i.e., AC 3, AC 2 or AC 1 can attempt

$0 \le Y_j^{(v)} \le N_v$ and $0 \le Y_j^{(t)} \le N_t$. Let $B_j^{(v)}$ be the number of new VoIP packet arrivals at all the $QSTA_v$s, in the channel slot j. Then $B_j^{(v)}$ is the number of $QSTA_v$s that add up for channel contention in the $(j+1)$th channel slot. Let $V_j^{(vAP)}$ be the number of packet departures from QAP_v, $V_j^{(vSTA)}$ be the number of departures from $QSTA_v$s, $V_j^{(vd)}$ be the number of departures from QAP_{vd}, $V_j^{(tAP)}$ be the number of departures from QAP_t and $V_j^{(tSTA)}$ be the number of departures from $QSTA_t$s, in the jth channel slot. We know that at most one departure can happen in any channel slot.

Then we have the following dynamics for the number of contending $QSTA$s.

$$Y_{j+1}^{(v)} = Y_j^{(v)} - V_{j+1}^{(vSTA)} + B_{j+1}^{(v)} \tag{1}$$

$$Y_{j+1}^{(t)} = Y_j^{(t)} - V_{j+1}^{(tSTA)} + V_{j+1}^{(tAP)} \tag{2}$$

with the condition: $V_{j+1}^{(vSTA)} + V_{j+1}^{(vAP)} + V_{j+1}^{(vd)} + V_{j+1}^{(tSTA)} + V_{j+1}^{(tAP)} \in \{0,1\}$, since, at most one node can succeed. Since the probability with which a packet arrives at a node in a channel slot of length l is p_l and we assume that packets arrive at only empty $QSTA_v$s, $B_j^{(v)}$ can be modeled using p_l (defined in Sect. 2.2) and the conditioned probability $Pr\left(B_{j+1}^{(v)} | (Y_j^{(v)}, L_{j+1}) = (n_v, l)\right)$ is given by

$$\begin{aligned}\Pr&\left(B_{j+1}^{(v)} = b \middle| \left(Y_j^{(v)} = n_v; L_{j+1} = l\right)\right)\\ &= \binom{N_v - n_v}{b} (p_l)^b (1-p_l)^{N_v - n_v - b}\end{aligned} \tag{3}$$

.

In the next sub-section we will make an approximation that permits us to determine expressions for $V_{j+1}^{(vSTA)}, V_{j+1}^{(vAP)}, V_{j+1}^{(vd)}, V_{j+1}^{(tSTA)}$ and $V_{j+1}^{(tAP)}$, and hence model the above dynamics (Eqs. 1 and 2) as a Markov chain embedded at channel slot boundaries.

3.2 Markov property via state dependent attempt probabilities

For determining the expressions of $V_{j+1}^{(vSTA)}, V_{j+1}^{(vAP)}, V_{j+1}^{(vd)}$, $V_{j+1}^{(tSTA)}$ and $V_{j+1}^{(tAP)}$, we need the attempt probabilities which we approximate as those obtained from the saturation results in [9]. But the AC attempt probabilities obtained from [9] are conditioned on when an AC can attempt. Note that after a channel activity, AC 1 cannot attempt and waits for an additional idle slot. We use the variable C_j to keep track of which ACs are permitted to attempt in a channel slot. Let $C_j = 1$ denote that the preceding channel slot had an activity and so in the beginning of the jth channel slot, only nodes with AC 3 or AC 2 can attempt. Let $C_j = 0$ denote that the preceding channel slot remained idle and hence, at the beginning of the jth channel slot any node can attempt. Thus $C_j \in \{0,1\}$.

In our model, if there are n_v non-empty $QSTA_v$s and n_t non-empty $QSTA_t$s, we have $n_v + 1$ AC 3 contending nodes, 1 AC 2 contending node and $n_t + 1$ AC 1 contending nodes, since QAP_v, QAP_{vd} and QAP_t, by assumption, are always non-empty. Let $\beta_{n_v+1,1,n_t+1}^{(v)}$ be the attempt probability of a AC 3 node, $\beta_{n_v+1,1,n_t+1}^{(vd)}$ be the attempt probability of a AC 2 node and $\beta_{n_v+1,1,n_t+1}^{(t)}$ be the attempt probability of a AC 1 node, when the nodes are non-empty. These attempt probabilities are conditioned on the event that the ACs can attempt. The values, $\beta_{n_v+1,1,n_t+1}^{(v)}, \beta_{n_v+1,1,n_t+1}^{(vd)}$ and $\beta_{n_v+1,n_t+1}^{(t)}$ are obtained from saturation fixed point analysis of [9] for all combinations of n_v,1,n_t. Our approximation is to use the state dependent values of attempt probabilities from the saturated nodes case, by keeping track of the number of nonempty nodes in the WLAN and whether the nodes can attempt, and taking the state dependent attempt probabilities corresponding to this number of nonempty nodes.

$$V_{j+1}^{(vSTA)} = \begin{cases} 1 & \text{w.p. } \alpha_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) \quad \text{if } C_j = 0 \\ 1 & \text{w.p. } \alpha_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) \quad \text{if } C_j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$V_{j+1}^{(vAP)} = \begin{cases} 1 & \text{w.p. } \sigma_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 0 \\ 1 & \text{w.p. } \sigma_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$V_{j+1}^{(vd)} = \begin{cases} 1 & \text{w.p. } \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_v(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 0 \\ 1 & \text{w.p. } \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})\eta_v(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$V_{j+1}^{(tSTA)} = \begin{cases} 1 & \text{w.p. } \alpha_t(Y_j^{(v)}, Y_j^{(t)})\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$V_{j+1}^{(tAP)} = \begin{cases} 1 & \text{w.p. } \sigma_t(Y_j^{(v)}, Y_j^{(t)})\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) & \text{if } C_j = 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

We use the thus obtained state dependent attempt probabilities to derive the probabilities of different activities in the channel. For convenience, let us define the following probability functions depicting the activities in the channel slot $j + 1$:

- $\eta_v(Y_j^{(v)}, Y_j^{(t)})$ be the probability that all nodes with AC 3 remain idle
- $\eta_t(Y_j^{(v)}, Y_j^{(t)})$ be the probability that all nodes with AC 1 remain idle
- $\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that QAP_{vd} remains idle
- $\alpha_v(Y_j^{(v)}, Y_j^{(t)})$ be the probability that exactly one $QSTA_v$ attempts while QAP_v is idle
- $\alpha_t(Y_j^{(v)}, Y_j^{(t)})$ be the probability that exactly one $QSTA_t$ attempts while QAP_t is idle
- $\sigma_v(Y_j^{(v)}, Y_j^{(t)})$ be the probability that the QAP_v attempts and all $QSTA_v$s are idle
- $\sigma_t(Y_j^{(v)}, Y_j^{(t)})$ be the probability that the QAP_t attempts and all $QSTA_t$s are idle
- $\sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that the QAP_{vd} attempts
- $\zeta_v(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a collision amongst AC 3 nodes (including QAP_v)
- $\zeta_t(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a collision amongst $QSTA_t$s
- $\psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a hybrid collision (collision between dissimilar packets) involving nodes with AC 3 (including QAP_v) and $QSTA_t$s
- $\psi_{v-vd}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a hybrid collision involving AC 3 nodes (including QAP_v) and QAP_{vd}
- $\psi_{vdAP}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a hybrid collision between QAP_{vd} and any other node, except QAP_t
- $\psi_{tAP}(Y_j^{(v)}, Y_j^{(t)})$ be the probability that there is a hybrid collision between QAP_t and any other node

The expressions for these functions are provided in Appendix A. We can then express the conditional distributions $V_{j+1}^{(vSTA)}, V_{j+1}^{(vAP)}, V_{j+1}^{(vd)}, V_{j+1}^{(tSTA)}$ and $V_{j+1}^{(tAP)}$ as follows: $V_{j+1}^{(vSTA)}$ is 1 if a $QSTA_v$ wins the contention for the channel and 0 otherwise, and is given by Eq. 4. Similarly $V_{j+1}^{(vAP)}, V_{j+1}^{(vd)}, V_{j+1}^{(tSTA)}$ and $V_{j+1}^{(tAP)}$ are given by Eqs. 5–8.

C_{j+1} takes the values in {0,1} with the following probabilities:

$$C_{j+1} = \begin{cases} 0 & \text{w.p. } \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) \\ 1 & \text{otherwise} \end{cases}$$

with the initial state, $C_0 = 0$.

With the assumed distribution for voice packet arrivals and the state dependent probabilities of attempt, it is easily seen from Eqs. 1 and 2 that $\{Y_j^{(v)}, Y_j^{(t)}, C_j; j \geq 0\}$ forms a finite irreducible three dimensional discrete time Markov chain on the channel slot boundaries and hence is positive recurrent. If n_v, n_t and c denote the sample variables of the random processes $Y_j^{(v)}, Y_j^{(t)}$ and C_j, respectively, the stationary probabilities $\pi_{n_v,n_t,c}$ of the Markov Chain $\{Y_j^{(v)}, Y_j^{(t)}, C_j; j \geq 0\}$ can be numerically determined (see Appendix B for details) using expressions of conditional distributions of $B_j^{(v)}$, and the probability functions expressed before.

3.3 The Markov renewal process

In this subsection we use the state dependent attempt probabilities to obtain the distribution of the channel slot duration. On combining this with the Markov chain in

Sect. 3.2, we finally conclude that $\{(Y_j^{(v)}, Y_j^{(t)}, C_j; U_j); j \geq 1\}$ is a Markov renewal process.

We use the basic access mechanism[1] for the channel access of all ACs. This shall facilitate the validation of analytical results through simulations by the *ns-2* with EDCA implementation [24], that supports only basic access mechanism and not RTS/CTS mechanism. However, our analysis can be worked out for the RTS/CTS mechanism as well.[2]

When the basic access mechanism is used, values of $L_j; j \geq 0$ are obtained as follows. There are four different time lengths of collisions. The longest collision time is seen when a QAP_t packet collides with a packet of any other node. The next longer collision time is seen when QAP_{vd} packet collides with a packet of any other node, except QAP_t. A smaller collision time is seen when a VoIP packet collides with a packet of any other node except with a packet of QAP_t or QAP_{vd}. The shortest collision time is seen when only packets of $QSTA_t$s collide. Then L_j (in system slots) takes one of the nine values: 1 if it is an idle slot; T_{s-v} if it corresponds to a successful transmission of a AC 3 node; T_{s-tAP} if it corresponds to a successful transmission of QAP_t; T_{s-vdAP} if it corresponds to a successful transmission of a AC 2 node; T_{s-tSTA} if it corresponds to a successful transmission of $QSTA_t$; $T_{c-short}$ if it corresponds to a collision between $QSTA_t$s; $T_{c-voice}$ if it corresponds to a collision amongst nodes with AC 3 or between AC 3 nodes and any $QSTA_t$; T_{c-vd} if it corresponds to a collision between QAP_{vd} and any other node, except QAP_t; and T_{c-long} if it corresponds to a collision between QAP_t and any other node.

The various values of L_j (in seconds) are as follows:

- $$T_{s-v} = T_P + T_{PHY} + \frac{L_{MAC} + L_{voice}}{C_d} + T_{SIFS} + T_P + T_{PHY} + \frac{L_{ACK}}{C_c} + T_{AIFS(3)};$$
- $$T_{s-tAP} = T_P + T_{PHY} + \frac{L_{MAC} + L_{IPH} + L_{TCPH} + L_{data}}{C_d} + T_{SIFS} + T_P + T_{PHY} + \frac{L_{ACK}}{C_c} + T_{AIFS(1)};$$
- $$T_{s-vdAP} = T_P + T_{PHY} + \frac{L_{MAC} + L_{IPH} + L_{UDPH} + L_{video}}{C_d} + T_{SIFS} + T_P + T_{PHY} + \frac{L_{ACK}}{C_c} + T_{AIFS(2)};$$
- $$T_{s-tSTA} = T_P + T_{PHY} + \frac{L_{MAC} + L_{IPH} + L_{TCPACK}}{C_d} + T_{SIFS} + T_P + T_{PHY} + \frac{L_{ACK}}{C_c} + T_{AIFS(1)};$$
- $$T_{c-short} = T_P + T_{PHY} + \frac{L_{MAC}+L_{IPH}+L_{TCPACK}}{C_d} + T'_{EIFS} + T_{AIFS(1)};$$
- $$T_{c-voice} = T_P + T_{PHY} + \frac{L_{MAC}+L_{voice}}{C_d} + T'_{EIFS} + T_{AIFS(3)};$$
- $$T_{c-vd} = T_P + T_{PHY} + \frac{L_{MAC}+L_{IPH}+L_{UDPH}+L_{video}}{C_d} + T'_{EIFS} + T_{AIFS(2)};$$
- $$T_{c-long} = T_P + T_{PHY} + \frac{L_{MAC}+L_{IPH}+L_{TCPH}+L_{data}}{C_d} + T'_{EIFS} + T_{AIFS(1)};$$
- $$T'_{EIFS} = T_P + T_{PHY} + \frac{L_{ACK}}{C_c} + T_{SIFS}.$$

See Table 2 for the meaning and values of various parameters. The probability mass function of the channel slot duration L_j, for above values, can be worked out using the probability functions of Subsection 3.3 and the expression for mean cycle time $\mathsf{E}L_{j+1}$ is given in Appendix C. Let $\underline{Y}_j = \left(Y_j^{(v)}, Y_j^{(t)}, C_j\right)$ denote the state vector at the channel slot boundary U_j. Then we observe Eq. 9 and so conclude that $\{(Y_j^{(v)}, Y_j^{(t)}, C_j; U_j), j \geq 0\}$ is a Markov

Table 2 Parameters used in analysis and simulation for EDCA 802.11e WLAN

Parameter	Symbol	Value
PHY data rate	C_d	11 Mbps
Control rate	C_c	2 Mbps
G711 pkt size	L_{voice}	200 Bytes
Videostreaming pkt size	L_{video}	1,500 Bytes
Data pkt size	L_{data}	1,500 Bytes
TCP header size	L_{TCPH}	20 Bytes
TCP ACK pkt (header) size	L_{TCPACK}	20 Bytes
UDP header size	L_{UDPH}	20 Bytes
IP header size	L_{IPH}	20 Bytes
MAC Header size	L_{MAC}	288 bits
MAC-layer ACK Pkt Size	L_{ACK}	112 bits
PLCP preamble time	T_P	144 µs
PHY Header time	T_{PHY}	48 µs
AIFS(3) time	$T_{AIFS(3)}$	50 µs
AIFS(2) time	$T_{AIFS(2)}$	50 µs
AIFS(1) time	$T_{AIFS(1)}$	70 µs
SIFS time	T_{SIFS}	10 µs
CW_{min} for AC(3)		7
CW_{max} for AC(3)		15
CW_{min} for AC(2)		15
CW_{max} for AC(2)		31
CW_{min} for AC(1)		31
CW_{max} for AC(1)		1,023
Idle/system slot (802.11b)	δ	20 µs

[1] The basic access mechanism is one of the two access mechanisms based on the CSMA/CA (carrier sense multiple access/collision avoidance) protocol for wireless transmissions. The other is the RTS/CTS (request to send/ clear to send) mechanism. See [23] for details.

[2] The only change will be the values of various possible channel slot lengths, $L_j; j \geq 0$, due to the differences in packet transmission times.

renewal process with $L_j = U_j - U_{j-1}$ being the renewal cycle time.

$$\begin{aligned} & Pr\Big(\underline{Y}_{j+1} = \underline{y}, (U_{j+1} - U_j) \leq l | (\underline{Y}_0 = \underline{y}_0, \underline{U}_0 = \underline{u}_0), \\ & \quad (\underline{Y}_1 = \underline{y}_1, \underline{U}_1 = \underline{u}_1), \ldots, (\underline{Y}_j = \underline{y}_j, \underline{U}_j = \underline{u}_j)\Big) \\ & \quad = Pr\Big(\underline{Y}_{j+1} = \underline{y}, (U_{j+1} - U_j) \leq l | (\underline{Y}_j = \underline{y}_j, \underline{U}_j = \underline{u}_j)\Big) \end{aligned} \tag{9}$$

4 Obtaining performance measures

4.1 VoIP call capacity

Let A_j be the "reward" when the QAP_v wins the channel contention in jth channel slot, i.e., $[U_{j-1}, U_j)$. If $Y_{j-1}^{(v)} = n_v, Y_{j-1}^{(t)} = n_t$ and $C_{j-1} = c$ then we have,

$$A_j = \begin{cases} 1 \text{ w.p. } \sigma_v(n_v, n_t)\eta_t(n_v, n_t)\eta_{vd}(n_v, n_t) & \text{if } c = 0 \\ 1 \text{ w.p. } \sigma_v(n_v, n_t)\eta_{vd}(n_v, n_t) & \text{if } c = 1 \\ 0 \text{ otherwise} \end{cases}$$

Let $A(t)$ denote the cumulative reward until time t. Applying Markov regenerative analysis [19] we obtain the service rate of the AP, $\Theta_{AP-voip}(N_v, N_t)$, as given by

$$\Theta_{AP-voip}(N_v, N_t) = \lim_{t\to\infty} \frac{A(t)}{t} \overset{a.s.}{=} \frac{\sum_{n_v=0}^{N_v} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \pi_{n_v,n_t,c} \mathsf{E}_{n_v,n_t,c} A}{\sum_{n_v=0}^{N_v} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \pi_{n_v,n_t,c} \mathsf{E}_{n_v,n_t,c} L} \tag{10}$$

where, $\mathsf{E}_{n_v,n_t,c} A = \mathsf{E}\Big(A_j | (Y_{j-1}^{(v)}, Y_{j-1}^{(t)}, Y_{j-1}^{(s)}) = (n_v, n_t, c)\Big)$, $\mathsf{E}_{n_v,n_t,c} L = \mathsf{E}\Big(L_j | (Y_{j-1}^{(v)}, Y_{j-1}^{(t)}, Y_{j-1}^{(s)}) = (n_v, n_t, c)\Big)$, $\mathsf{E}_{n_v,n_t,c} L = \mathsf{E}(L_j | (Y_{j-1}^{(v)}, Y_{j-1}^{(t)}, Y_{j-1}^{(s)}) = (n_v, n_t, c))$ and $\Theta_{AP-voip}$ is in packets per slot.

Since the rate at which a single call sends data to the QAP_v is λ, and the QAP_v serves N_v such calls, the total arrival rate to the QAP_v is $N_v\lambda$. This rate should be less than $\Theta_{AP-voip}(N_v, N_t)$ for stability. Thus, a permissible combination of N_v VoIP calls and N_t TCP sessions, with QAP_{vd} saturated, while meeting the delay QoS of VoIP calls, must satisfy

$$\Theta_{AP-voip}(N_v, N_t) > N_v\lambda \tag{11}$$

The above inequality defines an outer bound on the admission region for VoIP. Note that we are asserting that the N_v that satisfies Inequality (11) also ensures the delay QoS. This is based on the observation in earlier research ([25] and [26]) that when the arrival rate is less than the saturation throughput then the delay is very small. We validate this approach by our simulation results in Sect. 6.

Remark The model discussed above does not give the video and TCP download throughput. This is due to our assumption that the voice queue of the QAP is saturated all the time. But actually, the voice queue of QAP saturates only at system capacity [20]. Thus if we follow the above method to obtain analytical video and TCP download throughput, we obtain under estimations of the throughputs. This problem can be solved by modeling the occupancies of QAP_v, which we carry out in the following subsection.

4.2 Streaming video and TCP download throughput

Depending on whether the QAP_v contains a packet, the total number of nonempty AC 3 nodes will be $Y_j^{(v)}$ (in case no packet is there in QAP_v) or $Y_j^{(v)} + 1$ (if QAP_v has at least one packet). We then need to know the state of the QAP_v so as to know the number of nonempty AC 3 nodes, at the channel slot boundaries. Therefore, we introduce another variable to track the number of packets in the QAP_v.

Let $X_j^{(v)}$ be the number of packets in the QAP_v and $B_j^{(vAP)}$ be the number of new packets arriving at the QAP_v at the end of jth channel slot. Then, the set of evolution equations are:

$$\begin{aligned} Y_{j+1}^{(v)} &= Y_j^{(v)} - V_{j+1}^{(vSTA)} + B_{j+1}^{(v)} \\ Y_{j+1}^{(t)} &= Y_j^{(t)} - V_{j+1}^{(tSTA)} + V_{j+1}^{(tAP)} \\ X_{j+1}^{(v)} &= X_j^{(v)} - V_{j+1}^{(vAP)} + B_{j+1}^{(vAP)} \end{aligned}$$

with the condition: $V_{j+1}^{(vSTA)} + V_{j+1}^{(vAP)} + V_{j+1}^{(vd)} + V_{j+1}^{(tSTA)} + V_{j+1}^{(tAP)} \in \{0, 1\}$, since, at most one node can succeed.

The expression for $B_j^{(vAP)}$ can be written on similar lines as $B_j^{(v)}$. Observe that if x packets are already there in QAP_v queue, at most $N_v - x$ packets can arrive before the QoS delay bound of the earliest arrived packet gets exceeded. Using the earlier definition of p_l, the conditional probability $Pr(B_{j+1}^{(vAP)} | X_j^{(v)}, L_{j+1})$ is given by

$$\begin{aligned} & \Pr\Big(B_{j+1}^{(vAP)} = b | (X_j^{(v)} = x; L_{j+1} = l)\Big) \\ & \quad = \binom{N_v - x}{b} (p_l)^b (1 - p_l)^{N_v - x - b} \end{aligned} \tag{12}$$

In order to take into account the fact that QAP_v may or may not be contending at any channel slot boundary, define $Z_j^{(v)} := Y_j^{(v)} + 1$ if $X_j^{(v)} \neq 0$ and $Z_j^{(v)} := Y_j^{(v)}$ if $X_j^{(v)} = 0$. Then the probability functions in Subsection 3.2 need a modification. Instead of $\beta_{Y_j^{(v)}+1, 1, Y_j^{(t)}+1}$, we now have to use $\beta_{Z_j^{(v)}, 1, Y_j^{(t)}+1}$.

We again see that, under our model for the attempt probabilities, $\{Z_j^{(v)}, Y_j^{(t)}, C_j, X_j^{(v)}; j \geq 0\}$ forms a finite state irreducible four dimensional discrete time Markov chain on the channel slot boundaries and hence is positive recurrent. The stationary probabilities $\pi_{n_v,n_t,c,x}$ can be numerically obtained.

Streaming Video Throughput: Let T_j be the reward when the QAP_{vd} wins the channel contention in jth channel slot. If $Z_{j-1}^{(v)} = n_v, Y_{j-1}^{(t)} = n_t$ and $C_{j-1} = c$, then we have,

$$T_j = \begin{cases} 1 \text{ w.p. } \sigma_{vd}(n_v, n_t)\eta_v(n_v, n_t)\eta_t(n_v, n_t) & \text{if } c = 0 \\ 1 \text{ w.p. } \sigma_{vd}(n_v, n_t)\eta_v(n_v, n_t) & \text{if } c = 1 \\ 0 \text{ otherwise} \end{cases}$$

Let $T(t)$ denote the cumulative reward of the QAP_t until time t. Again, applying Markov regenerative analysis [19], the video streaming throughput $\Theta_{AP-vd}(N_v, N_t)$ is given by Eq. 13.

TCP Download Throughput: Let R_j be the reward when the QAP_t wins the channel contention in jth channel slot. If $Z_{j-1}^{(v)} = n_v, Y_{j-1}^{(t)} = n_t$ and $C_{j\text{-}1} = c$, then we have,

$$R_j = \begin{cases} 1 \text{ w.p. } \sigma_t(n_v, n_t)\eta_v(n_v, n_t)\eta_{vd}(n_v, n_t) & \text{if } c = 0 \\ 0 \text{ otherwise} \end{cases}$$

Let $R(t)$ denote the cumulative reward of the QAP_t until time t. Again, applying Markov regenerative analysis [19], the TCP download throughput $\Theta_{AP-TCP}(N_v, N_t)$ is given by Eq. 14.

channel slot, with $W_j = 1$ if the channel slot activity is a video success and $W_j \neq 1$ for all other activities. See Fig. 3. Then, L_j being the length of the jth channel slot, we obtain

$$\begin{aligned} &Pr(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, L_{j+1} \le u | \boldsymbol{Y}_j = \boldsymbol{x}) = \\ &\quad Pr\big(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, L_{j+1} \le u, W_{j+1} \neq 1 | \boldsymbol{Y}_j = \boldsymbol{x}\big) \\ &\qquad + Pr\big(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, L_{j+1} \le u, W_{j+1} = 1 | \boldsymbol{Y}_j = \boldsymbol{x}\big) \end{aligned}$$

Let $q_{\boldsymbol{x}}(\boldsymbol{y}, w) = Pr\big(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, W_{j+1} = w | \boldsymbol{Y}_j = \boldsymbol{x}\big)$, where w indicates the activity. Then,

$$\begin{aligned} &Pr(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, L_{j+1} \le u, W_{j+1} = 1 | \boldsymbol{Y}_j = \boldsymbol{x}) \\ &= q_{\boldsymbol{x}}(\boldsymbol{y}, 1) Pr(L_{j+1} \le u | W_{j+1} = 1, \boldsymbol{Y}_j = \boldsymbol{x}, \boldsymbol{Y}_{j+1} = \boldsymbol{y}), \\ &Pr(\boldsymbol{Y}_{j+1} = \boldsymbol{y}, \boldsymbol{L}_{j+1} \le \boldsymbol{u}, \boldsymbol{W}_{j+1} \neq \boldsymbol{1} | \boldsymbol{Y}_j = \boldsymbol{x}) \\ &= \sum_{\forall w, w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{y}, w) Pr(L_{j+1} \le u | W_{j+1} = w, \boldsymbol{Y}_j = \boldsymbol{x}, \boldsymbol{Y}_{j+1} = \boldsymbol{y}) \end{aligned}$$

$$\Theta_{AP-vd}(N_v, N_t) = \lim_{t\to\infty} \frac{T(t)}{t} \overset{a.s.}{=} \frac{L_{video} \sum_{n_v=0}^{N_v+1} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \sum_{x=0}^{N_v} \pi_{n_v,n_t,c,x} \mathsf{E}_{n_v,n_t,c,x} T}{\delta \sum_{n_v=0}^{N_v+1} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \sum_{x=0}^{N_v} \pi_{n_v,n_t,c,x} \mathsf{E}_{n_v,n_t,c,x} L} \tag{13}$$

$$\Theta_{AP-TCP}(N_v, N_t) = \lim_{t\to\infty} \frac{R(t)}{t} \overset{a.s.}{=} \frac{L_{data} \sum_{n_v=0}^{N_v+1} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \sum_{x=0}^{N_v} \pi_{n_v,n_t,c,x} \mathsf{E}_{n_v,n_t,c,x} R}{\delta \sum_{n_v=0}^{N_v+1} \sum_{n_t=0}^{N_t} \sum_{c=0}^{1} \sum_{x=0}^{N_v} \pi_{n_v,n_t,c,x} \mathsf{E}_{n_v,n_t,c,x} L} \tag{14}$$

where, $\mathsf{E}_{n_v,n_t,c,x} T(R) = \mathsf{E}(T_j(R_j) | (Z_{j-1}^{(v)}, Y_{j-1}^{(t)}, C_{j-1}, X_{j-1}^{(v)}) = (n_v, n_t, c, x))$, $\mathsf{E}_{n_v,n_t,c,x} L = \mathsf{E}(L_j | (Z_{j-1}^{(v)}, Y_{j-1}^{(t)}, C_{j-1}, X_{j-1}^{(v)}) = (n_v, n_t, c, x))$; Θ_{AP-vd} and Θ_{AP-TCP} are in Bps.

5 Further analysis of streaming video

5.1 Distribution of video service time

In this section we obtain the Laplace-Stieltjes transform (LST) of the video packet service time distribution at QAP_{vd} when the queue is saturated. This can then be used to obtain the maximum video throughput and provides an alternative method.

Let the sequence of random variables, $\{H_i,\ i \ge 1\}$ denote the service times of video packets (including the time of transmission of the video packet) when the QAP_{vd} is *saturated*. See Fig. 3. We denote the channel slot boundaries that end with a video packet success by $U_{j_k}, k \ge 1$, where k denotes the kth video packet success; for example, in Fig. 3, $j_1 = 3, j_2 = 7$, etc. Letting $j_0 = 0$, $H_i = U_{j_i} - U_{j_{(i-1)}}$. Let $H(\cdot)$ be the stationary distribution of $\{H_i,\ i \ge 1\}$ and denote the LST of $H(\cdot)$ by $\tilde{h}(s)$.

Let $\boldsymbol{Y}_j = \left(Z_j^{(v)}, Y_j^{(t)}, C_j, X_j^{(v)}\right)$ denote the state vector at the channel slot boundary U_j. Let χ be the set of all possible state vectors. Let W_j denote the type of activity in the jth channel slot, with $W_j = 1$ if the channel slot activity is a video success and $W_j \neq 1$ for all other activities.

Define $Pr\big(L_{j+1} \le u | W_{j+1} = w, \boldsymbol{Y}_j = \boldsymbol{x}, \boldsymbol{Y}_{j+1} = \boldsymbol{y}\big) := L_{\boldsymbol{xy},w}(u)$ and let its LST be $\tilde{l}_{\boldsymbol{xy},w}(s)$. $L_{\boldsymbol{xy},w}(u)$ is the distribution of the channel slot duration given the states at the two end points of the channel slot and the activity in the slot.

Consider a channel slot boundary U_j with $\boldsymbol{Y}_j = \boldsymbol{x}$. Let $G_{\boldsymbol{x}}$ be the random variable that denotes the time until the next video packet success is complete, starting with state $\boldsymbol{x}$. Let $G_{\boldsymbol{x}}(\cdot)$ denote its distribution and $\tilde{g}_{\boldsymbol{x}}(s)$ denote its LST. Then

$$\begin{aligned} \tilde{g}_{\boldsymbol{x}}(s) &= \sum_{\boldsymbol{y} \in \chi} q_{\boldsymbol{x}}(\boldsymbol{y}, 1) \tilde{l}_{\boldsymbol{xy},1}(s) \\ &\quad + \sum_{\boldsymbol{y} \in \chi} \left(\sum_{\forall w, w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{y}, w) \tilde{l}_{\boldsymbol{xy},w}(s) \right) \tilde{g}_{\boldsymbol{y}}(s) \end{aligned} \tag{15}$$

The first term in the above expression is for when there is a video packet success in the next channel slot. The second term is for the case when there is some other activity in the next channel slot and the slot ends in state $\boldsymbol{y}$; hence the term $\tilde{g}_{\boldsymbol{y}}(s)$ is for the time-to-go until the video success.

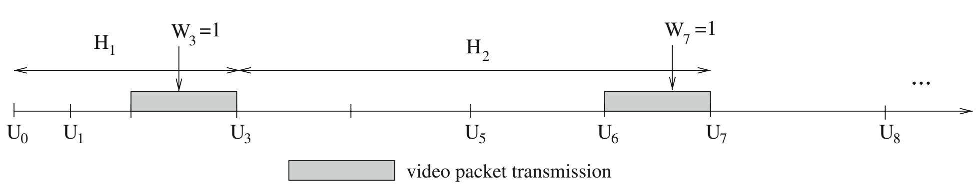

Fig. 3 The evolution activity of the channel showing the video packet success intervals, $H_j; j \in 0, 1, 2, \ldots$

Define $\{\boldsymbol{Y}_{j_k}, k \geq 1\}$ as the random process of state vectors at the boundaries of video packet success slots, i.e., at $U_{j_k}, k \geq 1$. We observe that $\{\boldsymbol{Y}_{j_k}, k \geq 1\}$ is also a finite irreducible Markov chain. Define $\boldsymbol{v}$ as the stationary probability vector over χ of this embedded Markov chain. Then $\widetilde{h}(s)$ can be expressed as

$$\widetilde{h}(s) = \sum_{\boldsymbol{x} \in \chi} v_{\boldsymbol{x}} \widetilde{g}_{\boldsymbol{x}}(s) \tag{16}$$

Now let $\widetilde{\boldsymbol{g}}(s)$ be the column vector with elements $\widetilde{g}_{\boldsymbol{x}}(s)$, $\boldsymbol{x} \in \chi$. Let $\boldsymbol{R}$ denote the $|\chi| \times |\chi|$ transition probability matrix with elements $q_{\boldsymbol{x}}(\boldsymbol{y}, 1)$. Let $\boldsymbol{Q}$ denote the matrix with elements $q_{\boldsymbol{x}}(\boldsymbol{y}, w) = \sum_{\forall w, w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{y}, w)$. Note that $\boldsymbol{R} + \boldsymbol{Q}$ forms a stochastic matrix. Let $\boldsymbol{Q}(s)$ denote the matrix with elements $q_{\boldsymbol{x}}(\boldsymbol{y}, w; s) = \sum_{\forall w, w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{y}, w) \widetilde{l}_{\boldsymbol{xy},w}(s)$. Let $\boldsymbol{1}$ be the column vector with all ones. Then Eq. 15 in matrix form is

$$\widetilde{\boldsymbol{g}}(s) = \boldsymbol{R1} e^{-sT_{s\text{-}vdAP}\delta} + \boldsymbol{Q}(s)\widetilde{\boldsymbol{g}}(s)$$

since $\widetilde{l}_{\boldsymbol{xy},1}(s) = e^{-sT_{s\text{-}vdAP}\delta}$. Here $T_{s\text{-}vdAP}$, is the time for successful transmission of a video packet, as defined earlier.

Solving the above equation for $\widetilde{\boldsymbol{g}}(s)$, we get

$$\widetilde{\boldsymbol{g}}(s) = (\boldsymbol{I} - \boldsymbol{Q}(s))^{-1} \boldsymbol{R1} e^{-sT_{s\text{-}vdAP}\delta} \tag{17}$$

The inverse $(\boldsymbol{I} - \boldsymbol{Q}(s))^{-1}$ can be shown to exist since $\boldsymbol{R} + \boldsymbol{Q}$ is irreducible and $\boldsymbol{R}$ is positive.

Equation 16 in matrix form is

$$\widetilde{h}(s) = \boldsymbol{v}\widetilde{\boldsymbol{g}}(s) \tag{18}$$

The stationary probability vector, $\boldsymbol{v}$ is obtained as follows: Let $\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R}$. Then

$$\boldsymbol{P} = \boldsymbol{R} + \boldsymbol{QR} + \boldsymbol{Q}^2\boldsymbol{R} + \boldsymbol{Q}^3\boldsymbol{R}\ldots$$

and we note that the $(\mathbf{x}, \mathbf{y})$ element of the kth term in the above expression corresponds to a video packet success at the kth channel slot, $k \geq 1$, with the initial state being $\mathbf{x}$ and the state just after the video success being $\mathbf{y}$. Thus $\boldsymbol{P}$ is the transition probability matrix for the Markov chain $\{\boldsymbol{Y}_{j_k}; k \geq 1\}$. Then $\boldsymbol{v} = \boldsymbol{vP}$ and we can numerically obtain $\boldsymbol{v}$.

The LST of video service time distribution can then be used to obtain the mean service time $\mathsf{E}H$, and hence the average video throughput, i.e., $\Theta_{AP-vd} = \frac{L_{video}}{\mathsf{E}H}$, where $\mathsf{E}H = -\frac{d}{ds}\widetilde{h}(s)\Big|_{s=0}$. The numerical values for Θ_{AP-vd} obtained this way tally with those obtained from Eq. 13, and further validate our analysis (see Fig. 6 for the values of Θ_{AP-vd}, for different N_v).

5.2 Video packet loss and buffer sizing

Streaming video does not have any intrinsic delay objective, since the playout device can, in principle, compensate for substantial amounts of delay. However, the QAP_{vd} has a finite buffer. Hence, increasing the input video rate to values close to Θ_{AP-vd} will result in packet losses. Evidently, a large packet loss rate will not be tolerated by the video decoder and will result in poor video quality. It is thus of interest to study the video packet loss probability in order to size the QAP_{vd} buffer.

To obtain the size of the QAP_{vd} buffer to meet a given packet loss probability, we follow the well known approach of effective bandwidths (see [27, Chapter 5] and the references therein). The approach is based on an application of Chernoff's bound and on the log moment generating function of the arrival process.

Let the buffer size of QAP_{vd} be B (in packets). Consider the video packet loss probability constraint to be 'probability of packet loss $< \epsilon$'. We model the video packet arrival process into the AP video buffer as a Poisson process. This will be a good approximation if several video streams are multiplexed, and will yield a bound on B if we actually have one CBR video. Let us assume a total video packet arrival rate of λ_{vd}.

(a) Approximation via Level Crossing in an Infinite Buffer: Let $X^{(vd)}(t)$ denote the video buffer occupancy in the AP at time $t \geq 0$. Let $X_j^{(vd,a)}$ denote the process of the number of video packets seen by the jth video packet arrival, and let $X^{(vd,a)}$ denote its stationary random variable. With B finite, we are interested in the video packet loss probability, i.e.,

$$Pr(X^{(vd,a)} = B) = \lim_{t \to \infty} \frac{1}{\Lambda(t)} \sum_{j=1}^{\Lambda(t)} I_{\{X^{(vd)}(t_{j-}) = B\}}$$

where $t_j, j \geq 1$, denote the successive arrival instants of video packets, and $\Lambda(t)$ denotes the cumulative number of video packet arrivals until t. $I_{\{X^{(vd)}(t_{j-}) = B\}}$ is, as usual, an indicator function and $t_{j\text{-}}$ denotes that the arrival is not included.

Now, let $X^{(\infty)}(t)$ denote the video buffer process for an infinite buffer. Let, for $j \geq 1$, $X_j^{(\infty,a)} := X^{(\infty)}(t_{j-})$, i.e., $X_j^{(\infty,a)}$ is the number in the buffer "seen" by the jth video packet arrival (with infinite buffer). Further, let $X^{(\infty,a)}$ denote the stationary random variable for the process $X_j^{(\infty,a)}, j \geq 1$. Then $Pr(X^{(\infty,a)} > B - 1)$ will yield an upper bound on the desired probability $Pr(X^{(vd,a)} = B)$. Hence, in order to bound $Pr(X^{(vd,a)} = B)$ by ε we seek to achieve $Pr(X^{(\infty,a)} > B - 1) < \varepsilon$.

Let, with infinite buffer, $X_k^{(\infty,d)}, k \geq 1$, denote the number of video packets left behind by the kth video packet transmission. A standard rate balance argument (see [18]) then allows us to conclude that

$$Pr(X^{(\infty,a)} > B - 1) = Pr(X^{(\infty,d)} > B - 1) \tag{19}$$

From Eq. 19 we conclude that we need to study $Pr(X^{(\infty,d)} > B - 1)$, i.e., the stationary distribution of video packets at video packet transmission completion instants. To do this, we make one more approximation.

Whenever the video queue in the AP becomes empty, we insert a *dummy* video packet in the buffer. This ensures that the video queue in the AP is always contending and we can use the service process model in Sect. 5.1. If a video packet arrives while the dummy packet is contending, we replace the dummy packet with the arriving video packet. This simplification will provide a good approximation for video rates close to saturation and will yield a bound on the buffer required. We will require that $\lambda_{vd} < \frac{1}{\mathsf{E}H}$, with $\mathsf{E}H$ as defined in Sect. 5.1. We will call the service completion instants at the video queue in the AP, either of real video packets or dummy video packets, as *virtual* service instants.

Now we will make an argument that relates $Pr(X^{(\infty,d)} > b)$, for some b, to the distribution of the state at virtual service instants of the video queue at the AP. Let $S_k^{(\infty)}$ denote the number of video packets at the kth such virtual service instant (in the infinite buffer system). Let $\{\Lambda_k, k \geq 1\}$ denote the number of video packet arrivals in the time between the $(k-1)$th and kth virtual service instants. Then we observe that

$$S_k^{(\infty)} = \left(S_{k-1}^{(\infty)} + \Lambda_k - 1\right)^+ \tag{20}$$

The kth such service is that of a dummy packet iff $S_{k-1}^{(\infty)} + \Lambda_k = 0$. Define a sequence of random variables D_k, with $D_k = 1$ if a real video packet is served at the kth virtual service instant, and $D_k = 0$ otherwise. Then, we can see that, with probability one,

$$Pr(X^{(\infty,d)} > b) = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} I_{\{S_k^{(\infty)} > b, D_k = 1\}}}{\sum_{k=1}^{n} I_{\{D_k=1\}}} \tag{21}$$

For $b > 0$, it is clear that $I_{\{S_k^{(\infty)} > b, D_k=1\}} = I_{\{S_k^{(\infty)} > b\}}$. For the model in Eq. 20, we see that $\lambda_{vd} < \frac{1}{\mathsf{E}H}$ ensures that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_{\{S_k^{(\infty)} > b\}} = Pr(S^{(\infty)} > b) \tag{22}$$

where $S^{(\infty)}$ denotes the stationary random variable for the process $S_k^{(\infty)}$. Let $K(t)$ denote the number of virtual service completions until t. Then, $K(t) \to \infty$ with probability 1, and we observe that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_{\{D_k=1\}} &= \lim_{t\to\infty} \frac{1}{K(t)} \sum_{k=1}^{K(t)} I_{\{D_k=1\}} \\
&= \lim_{t\to\infty} \frac{t}{K(t)} \frac{1}{t} \sum_{k=1}^{K(t)} I_{\{D_k=1\}} \\
&= \mathsf{E}H \; \lambda_{vd} \\
&=: \rho(<1)
\end{aligned} \tag{23}$$

i.e., the fraction of virtual services that are real video packet services is $\rho = \mathsf{E}H \; \lambda_{vd}$. We conclude, from Eqs. 21–23, that

$$Pr(X^{(\infty,d)} > b) = \frac{Pr(S^{(\infty)} > b)}{\rho}$$

In particular, in order to ensure $Pr(X^{(\infty,d)} > B-1) < \epsilon$ we need to ensure that

$$Pr(S^{(\infty)} > B-1) < \rho\epsilon \tag{24}$$

(b) Using Chernoff's Bound: Thus, we wish to obtain $Pr(S^{(\infty)} > B-1) < \rho\varepsilon$, where $S^{(\infty)}$ is the stationary random variable for the stochastic recursion in Eq. 20. We follow the Chernoff bound based "effective bandwidth" approach (see [27, Chapter 5] and the references therein). Define

$$\Gamma(\theta) = \lim_{n\to\infty} \frac{1}{n} \ln \mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k} \tag{25}$$

for $\theta > 0$. Note that the distribution $\boldsymbol{v}$ is (as in Sect. 5.1) that of the state of the contending queues (other than the video queue at the AP) at the virtual service instants. Define $\theta = -\frac{\ln(\rho\epsilon)}{B-1}$. Then $Pr(S^{(\infty)} > B-1)$ is obtained if $\frac{\Gamma(\theta)}{\theta} < 1$, where the 1 is just the maximum amount by which $S_k^{(\infty)}$ is reduced by in each step of the recursion in Eq. 20. Note that the approximation will yield a bound on the required buffer. We will use simulations to study how loose this bound is.

We first calculate $\Gamma(\theta)$ as follows: $\mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k}$ can be split as

$$\mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k} = \mathsf{E}_{\boldsymbol{v}} e^{\theta \Lambda_1} e^{\theta \sum_{k=2}^{n} \Lambda_k} \tag{26}$$

Using the notation introduced in Sect. 5.1, let $p_{\boldsymbol{x}}(\boldsymbol{y})$ denote the elements of the transition matrix $\boldsymbol{P}$. Then we can continue the above equation as follows

$$= \sum_{\boldsymbol{x}\in\chi} v_{\boldsymbol{x}} \left(\sum_{\boldsymbol{y}\in\chi} p_{\boldsymbol{x}}(\boldsymbol{y}) \mathsf{E}_{\boldsymbol{x},\boldsymbol{y}} e^{\theta\Lambda_1} \right) \mathsf{E}_{\boldsymbol{y}} e^{\theta \sum_{k=1}^{n-1} \Lambda_k}$$

where $\mathsf{E}_{\boldsymbol{x},\boldsymbol{y}} e^{\theta\Lambda_1}$ is the moment generating function of Poisson arrivals in the time between two virtual service instants when the states at these two instants are $\boldsymbol{x}$ and $\boldsymbol{y}$.

Let us denote

$$\mu_{\boldsymbol{x}}(\boldsymbol{y}, \theta) = p_{\boldsymbol{x}}(\boldsymbol{y}) \mathsf{E}_{\boldsymbol{x},\boldsymbol{y}} e^{\theta\Lambda_1}$$

and $\mathsf{E}_{\boldsymbol{y}} e^{\theta \sum_{k=1}^{n-1} \Lambda_k} := f_{\boldsymbol{y}}(n-1, \theta)$, and let $\boldsymbol{M}(\theta)$ be the $|\chi| \times |\chi|$ matrix with elements $\mu_{\boldsymbol{x}}(\boldsymbol{y}, \theta)$. Let $\boldsymbol{f}(n-1, \theta)$ be the column vector with elements $f_{\boldsymbol{y}}(n-1, \theta)$ for all $\boldsymbol{y} \in \chi$. Then we can write

$$\mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k} = \boldsymbol{v}\boldsymbol{M}(\theta)\boldsymbol{f}(n-1, \theta) \tag{27}$$

Recursing this equation, we finally obtain

$$\mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k} = \boldsymbol{v}(\boldsymbol{M}(\theta))^{n-1}\boldsymbol{f}(1, \theta) \tag{28}$$

where $\boldsymbol{f}(1, \theta)$ is the column vector with the elements $f_{\boldsymbol{y}}(1, \theta)$. It remains to determine the matrix $\boldsymbol{M}(\theta)$ and the vector $\boldsymbol{f}(1, \theta)$.

(c) Analysis of $\boldsymbol{M}(\theta)$: As in Sect. 5.1, $w = 1$ denotes channel slot activity corresponding to a video packet success, and $w \neq 1$ correspond to other activities, such as voice packet success, TCP ACK packet collisions, etc. Then $\mu_{\boldsymbol{x}}(\boldsymbol{y}, \theta)$ can be obtained by conditioning on the kind of activity in the first channel slot. Let the channel slot length due to an activity w be $l(w)$. Then the m.g.f. of the number of Poisson arrivals in a slot with activity w is $e^{-\lambda_{vd} l(w)(1-e^{\theta})}$. Observing that, given the activity in a slot, the time taken by the activity is independent of the next state at the end of the slot, we can write

$$\mu_{\boldsymbol{x}}(\boldsymbol{y}, \theta) = \sum_{z \in \chi} \left(\sum_{w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{z}, w) e^{-\lambda_{vd} l(w)(1-e^{\theta})} \right) \mu_{\boldsymbol{z}}(\boldsymbol{y}, \theta) + q_{\boldsymbol{x}}(\boldsymbol{y}, 1)\ e^{-\lambda_{vd} l(1)(1-e^{\theta})} \quad (29)$$

where $q_{\boldsymbol{x}}(\boldsymbol{z}, w)$ and $q_{\boldsymbol{x}}(\boldsymbol{y}, 1)$ are as in Sect. 5.1.

Let $\boldsymbol{N}(\theta)$ denote the $|\chi| \times |\chi|$ matrix with elements $\sum_{w \neq 1} q_{\boldsymbol{x}}(\boldsymbol{z}, w) e^{-\lambda_{vd} l(w)(1-e^{\theta})}$ for all $\boldsymbol{x}$ and $\boldsymbol{z}$, and let $\boldsymbol{V}(\theta)$ be the $|\chi| \times |\chi|$ matrix with elements $q_{\boldsymbol{x}}(\boldsymbol{y}, 1) e^{-\lambda_{vd} l(1)(1-e^{\theta})}$. Then, Eq. 29 can be written in matrix form as

$$\boldsymbol{M}(\theta) = \boldsymbol{N}(\theta)\boldsymbol{M}(\theta) + \boldsymbol{V}(\theta) \quad (30)$$

(d) Analysis of $\boldsymbol{f}(1, \theta)$: It can also be seen that

$$\boldsymbol{f}(1, \theta) = \boldsymbol{N}(\theta)\boldsymbol{f}(1, \theta) + \boldsymbol{v}(\theta) \quad (31)$$

where $v_{\boldsymbol{y}}(\theta) = \sum_{z \in \chi} q_{\boldsymbol{y}}(\boldsymbol{z}, 1) e^{-\lambda_{vd} l(1)(1-e^{\theta})}$, with $q_{\boldsymbol{y}}(\boldsymbol{z}, 1)$ as defined in Sect. 5.1.

Theorem 5.1 *If θ is such that $\boldsymbol{M}(\theta)$ is a finite valued irreducible matrix, then $\Gamma(\theta) (= \lim_{n\to\infty} \frac{1}{n} \ln \mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k}) = \ln \xi(\theta)$, where $\xi(\theta)$ is the Perron–Frobenius eigenvalue of $\boldsymbol{M}(\theta)$.*

Proof We have from Eq. 28 that

$$\mathsf{E}_{\boldsymbol{v}} e^{\theta \sum_{k=1}^{n} \Lambda_k} = \boldsymbol{v}(\boldsymbol{M}(\theta))^{n-1} \boldsymbol{f}(1, \theta)$$

For finite $\boldsymbol{M}(\theta)$ we conclude from Eqs. 30 to 31 that $\boldsymbol{f}(1, \theta)$ is also finite, and then it follows from [28, Theorem 3.1.1] that

$$\lim_{n\to\infty} \frac{1}{n} \left(\ln \boldsymbol{v}(\boldsymbol{M}(\theta))^{n-1} \boldsymbol{f}(1, \theta) \right) = \ln \xi(\theta)$$

where $\xi(\theta)$ is the Perron–Frobenius eigenvalue of $\boldsymbol{M}(\theta)$.□

We observe that, since,

$$\mathsf{E}_{v} e^{\theta \Lambda_1} = \sum_{\boldsymbol{x} \in \chi} v_{\boldsymbol{x}} \left(\sum_{\boldsymbol{y} \in \chi} \mu_{\boldsymbol{x}}(\boldsymbol{y}, \theta) \right)$$

and χ is a finite set, $\boldsymbol{M}(\theta)$ is a finite matrix if and only if $\mathsf{E}_{v} e^{\theta \Lambda_1}$ is finite. We use this criterion to check the hypothesis of Theorem 5.1 in our numerical calculations below.

Thus, $\Gamma(\theta)$ in Eq. 25 is numerically calculated. We then plot $\frac{\Gamma(\theta)}{\theta}$ for various values of B, in order to determine the buffer size of QAP_{vd}. The results are provided in Sect. 6.4.

6 Numerical results and validation

We present the results obtained from the analysis and simulation. The simulations were obtained using *ns-2* with EDCA implementation [24]. VoIP traffic was considered on AC 3, video streaming traffic was considered on AC 2 and the TCP traffic was considered on AC 1. The PHY parameters conform to the 802.11b standard. See Table 2 for the values used in simulation.

In simulations, the start time of a VoIP call is uniformly distributed in [0, 20 ms]. This ensures that the voice packets do not arrive in bursts and remain non synchronized.

When the WLAN consists of only TCP download traffic, the analytical model for TCP download traffic is accurate for 5 or more TCP sessions (see [20] and [29]). Further, the analytical and simulation results confirmed that the aggregate download throughput is insensitive to the increase in the number of TCP sessions. In the present context where all kinds of traffic are present, the model again predicts accurate results for 5 or more TCP sessions and the results for $N_t > 5$ are same as for $N_t = 5$. Hence, in all cases of results, when TCP traffic is present, we consider $N_t = 5$.

For all numerical and simulation results, VoIP packet size is 200 bytes (G711 Codec); video stream packet size is 1,500 bytes; TCP data packet size is 1,500 bytes; PHY data rate is 11 Mbps and control rate is 2 Mbps. In the simulation results, the error bars denote the 95% confidence intervals.

6.1 VoIP capacity

In Fig. 4, we show the analytical plot of QAP_v service rate vs. the number of calls, N_v for cases when only VoIP calls are present and when VoIP calls are present along with video streaming and TCP download sessions. From Fig. 4, we note that the QAP_v service rate crosses the QAP_v load rate, after 12 calls for $N_t = 0$ and no video sessions. This implies that a maximum of 12 calls are possible while meeting the delay QoS, on a 802.11e WLAN when no other traffic is present. When video streaming sessions and TCP download sessions are also present in the WLAN, the QAP_v service rate crosses below the QAP_v load rate, after 7 calls. This implies that only 7 calls are possible when other traffics are present.

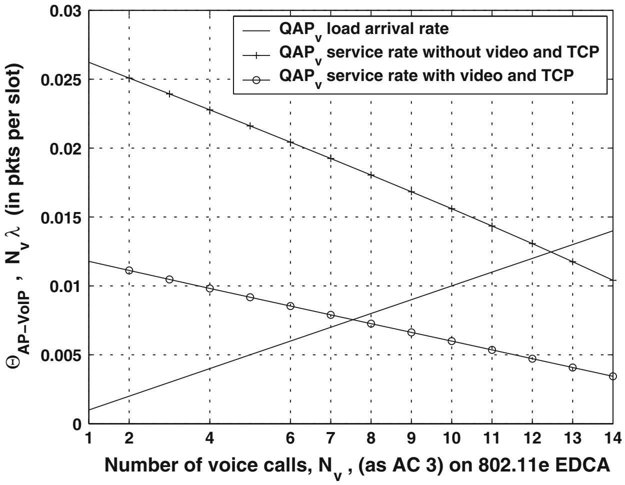

Fig. 4 The service rate $\Theta_{AP-voip}$ applied to the QAP_v is plotted as a function of number of voice calls, N_v, without and with video and TCP sessions. When present, the QAP_{vd} is assumed saturated and $N_t = 5$. Also shown is the line $N_v\ \lambda$. The point where the line $N_v\ \lambda$ crosses the curves gives the maximum number of calls supported

Remark The analysis represented by Fig. 4, assumes that the QAP_v is saturated. It is for this reason that the QAP_v service rate exceeds the load arrival rate for small N_v. The crossover point would however correctly model the value of N_v beyond which voice QoS will be violated.

Simulation results for the QoS objective of $Pr(delay \geq 20$ ms) for the QAP_v and the $QSTA_v$s are shown in Fig. 5. Note that the $Pr(delay{:}QAP_v \geq 20$ ms) is greater than $Pr(delay{:}QSTA_v \geq 20$ ms) for given N_v and that the QAP_v delay shoots up before the $QSTA_v$ delay, *confirming that the*QAP_v *is the bottleneck*, as per our assumptions. It

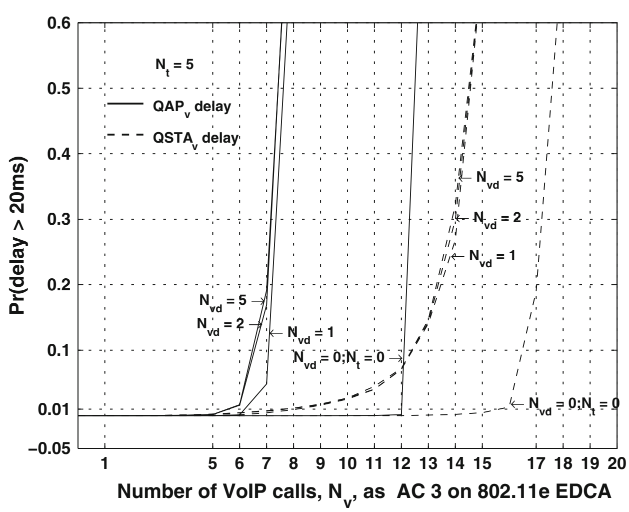

Fig. 5 Simulation results showing probability of delay of QAP_v and $QSTA_v$, being greater than 20 ms vs the number of calls (N_v) for different values of N_t. The solid lines denote the delay of QAP_v and the dashed lines denote the delay of $QSTA_v$

can be seen that with and without TCP traffic and video streaming traffic, there is a value of N_v at which the $Pr(delay{:}QAP_v \geq 20$ ms) sharply increases from a value below 0.01. This can be taken to be the voice capacity. When TCP and video traffic are present, we get a maximum of 6 calls, one less than the analysis result.

We have also done the analysis and simulations for the scenario when only VoIP and video streams are present in the WLAN (see [30]) and for the scenario when only VoIP and TCP downloads are present in the WLAN (see [17]). We summarize the results of all scenarios in Table 3.

6.2 Video throughput

We plot the analytical and simulation saturation throughput of video sessions vs the number of VoIP calls in Fig. 6. The number of TCP sessions, $N_t = 5$. The video sessions are assumed to be using 1,500 byte packets. The video queue of QAP in the simulation is saturated by sending a high input CBR traffic (more than 5 Mbps). We observe that the analytical results match very closely with the simulation results for different number of VoIP calls. For instance, for $N_v = 4$, the numerical saturation video throughput is 3.25 Mbps while the simulation value is 3.26 Mbps. Note that the plot after $N_v = 6$ calls is not of any use because, from Fig. 5 we already saw that the VoIP delay QoS breaks down after $N_v = 6$ calls. The error between the analysis and simulation then, is less than 5%, in the admissible region of VoIP calls. We note that a reduction of one VoIP call increases the video downlink stream throughput by approximately 0.38 Mbps.

We now consider the actual SD-TV quality video streaming sessions with a rate of 1.5 Mbps [22] between the server on the local network and the $QSTA_{vd}$s. This implies that the QAP_{vd} receives CBR video streams in multiples of 1.5 Mbps from the streaming server, as per the number of video streaming sessions. In Fig. 7 we plot the simulation results for the aggregate video streaming throughput obtained when the video streams are considered as CBR, with a rate of 1.5 Mbps and packet size of 1,500 bytes. Along side, the figure shows the saturation video throughput obtained from the analysis. The figure shows that as long as the available throughput (the saturation throughput) is above the required throughput, the video sessions obtain their required throughput. For instance, when two video streaming sessions are present, the total required throughput is 3 Mbps. We see that until $N_v = 4$, the video streams get an aggregate of 3 Mbps but when $N_v = 5$, the aggregate throughput is less than the required throughput. Note that at $N_v = 5$, the analytical saturation video throughput is 2.88 Mbps, which is less than the required throughput of 3 Mbps.

Table 3 Summary of VoIP capacity for an infrastructure 802.11e WLAN with EDCA

Max number of voice calls, N_{max}							
With out TCP and with out video		With TCP and with out video		With out TCP and with video		With TCP and with video	
Anal	Sim	Anal	Sim	Anal	Sim	Anal	Sim
12	12	10	9	8	8	7	6

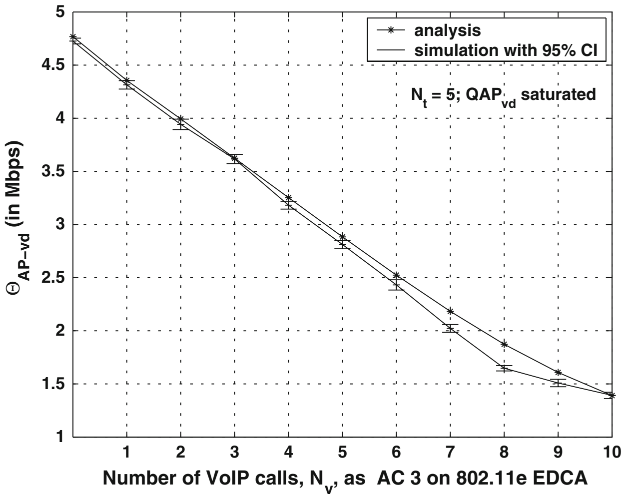

Fig. 6 Analysis and simulation results showing saturation video throughput Θ_{AP-vd} obtained by the QAP_{vd}, plotted as a function of number of voice calls, N_v

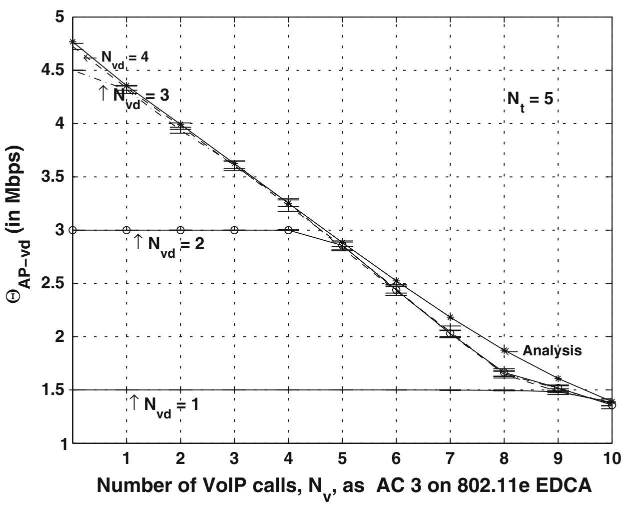

Fig. 7 Simulation results showing video throughput Θ_{AP-vd} obtained by the QAP_{vd}, plotted as a function of number of voice calls, N_v. The video streaming sessions are of 1.5 Mbps rate. The analytical saturation video throughput is shown alongwith for reference

6.3 TCP download throughput

The analytical and simulation results for aggregate TCP download throughput obtained by TCP sessions vs the number of VoIP calls is shown in Fig. 8. The number of

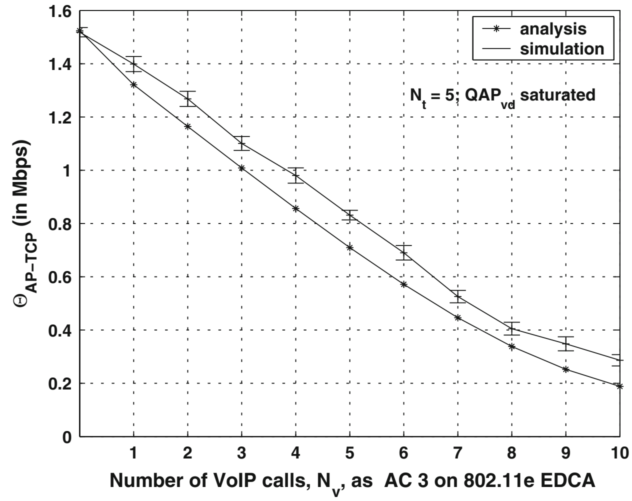

Fig. 8 Analysis and simulation results showing aggregate download throughput obtained by $QSTA_t$s for different values of N_v and $N_t = 5$, when QAP_{vd} is saturated

TCP sessions, $N_t = 5$. The video sessions are assumed to be using 1,500 bytes, with QAP_{vd} being saturated. For instance, for $N_v = 3$, the aggregate throughput obtained from analysis is 1.01 Mbps and that obtained from simulations is 1.10 Mbps.

We note that though the analytical curve follows the nature of the simulation curve, it underestimates the aggregate TCP throughput by at most 100 Kbps when compared with the simulations. Also, reducing the voice call by one increases the file download throughput by 0.14 Mbps approximately.

Figure 9 shows the simulation results of aggregate TCP download throughput when the QAP_{vd} is not saturated, but instead, the video sessions are CBR with packet size of 1,500 bytes and 1.5 Mbps rate. The figure shows the plots for different number of video sessions. The two curves at the bottom are same as shown in Fig. 8. The curves that start higher on the Θ_{TCP} axis and then drop to meet the curves of Fig. 8 correspond to 0, 1, 2 and 3 video streams. For $N_{vd} = 4$, the QAP_{vd} saturates and so coincides with the simulation curve of Fig. 8. As N_v increases, for each value of N_{vd}, the TCP throughput decreases until it meets the curves in Fig. 8.

Remark When the video sessions do not saturate the QAP_{vd}, more transmission opportunities are obtained by

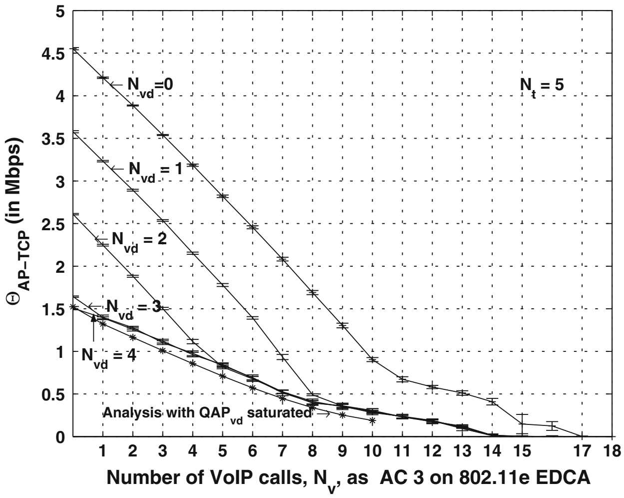

Fig. 9 Simulation results showing aggregate TCP download throughput obtained by $QSTA_t$s for different values of N_v and $N_t = 5$; The video streaming sessions are of 1.5 Mbps rate. The analytical aggregate TCP download throughput when QAP_{vd} is saturated, is shown alongwith for reference

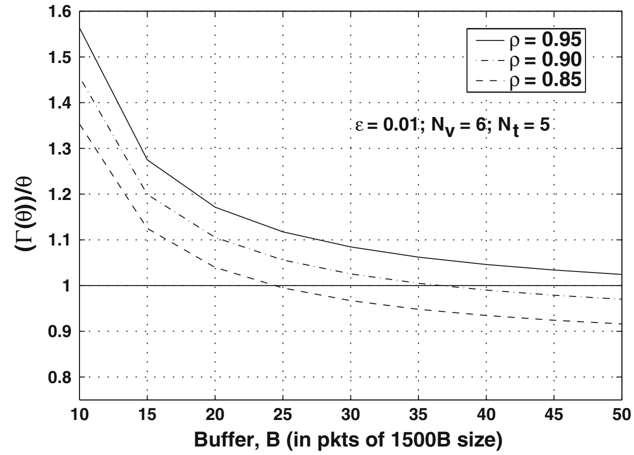

Fig. 10 Analysis results showing effective bandwidth $\frac{\Gamma(\theta)}{\theta}$ vs. B, for $\varepsilon = 0.01$.The three curves are for different ρ. $N_v = 6$ and $N_t = 5$

the TCP packets at QAP_t and hence the TCP aggregate throughput is more than that obtained when QAP_{vd} is saturated. For instance consider the curve when $N_{vd} = 2$. For $N_v = 2$, the simulation TCP throughput is 1.9 Mbps (see Fig. 9) against 1.3 Mbps (see Fig. 8), when QAP_{vd} is saturated. But however, after $N_v = 5$, the simulation curve follows the analytical curve. It can be noted that our analysis does not capture the performance of TCP traffic in the region when the video queue is not saturated. This is because in the model, we always consider a saturated QAP_{vd}. To obtain the TCP throughput when the video queue is not saturated, we need to model the video traffic also, which, due to varied codecs of use and different rates of encoding for desired quality of video streaming sessions, becomes complicated.

6.4 AP video buffer sizing

In this section we report numerical results based on the analysis developed in Sect. 5.2 and validate them with simulation results. We recall the definition: $\rho = \mathsf{E}H\ \lambda_{vd}$, which can be viewed as the load on QAP_{vd}, the AP video queue. In each case when we calculate $\Gamma(\theta)$, we have ensured that the matrix $\boldsymbol{M}(\theta)$ is finite via the observation following Theorem 5.1.

Figure 10 shows the analytical plot of $\frac{\Gamma(\theta)}{\theta}$ vs. B for $\varepsilon = 0.01$, when $N_v = 6$ and $N_t = 5$. Note that $N_v = 6$ corresponds to the maximum number of VoIP calls possible and hence leads to maximum buffer fill up at QAP_{vd}. We note that the curve corresponding to $\rho = 0.9$ cuts the $\frac{\Gamma(\theta)}{\theta} = 1$ line after $B = 37$. For $\rho = 0.85$, $\frac{\Gamma(\theta)}{\theta} < 1$ after $B = 24$. We can thus conclude from these analytical results that in the region of operation of traffic while meeting their QoS, the video streams can be guaranteed "probability of loss < 0.01", with about 40 packets buffer size at the QAP_{vd}.

We provide the simulation results in Fig. 11. In order to verify the analysis, we have considered Poisson arrivals at QAP_{vd} in the simulations. We observe from the figure that the video packet loss probability falls below 0.01 at $B = 28$, for $\rho = 0.90$, as compared to $B = 37$ obtained from the analysis (Fig. 10). For $\rho = 0.85$, we need $B = 17$ to ensure loss probability below 0.01, as compared to $B = 24$ from the analysis (Fig. 10). In both the cases, the required buffer sizes are less than obtained from the analysis. This is to be expected, since the analysis is based on a bound. This bound could be further improved by using a correction to the effective bandwidth based analysis (see [27, Chapter 5]).

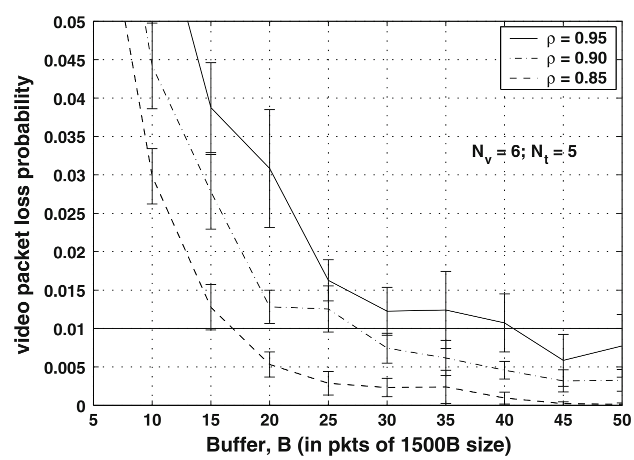

Fig. 11 Simulation results showing video packet loss probability vs. B, for $\varepsilon = 0.01$.The video packet arrival process is Poisson. $N_v = 6$ and $N_t = 5$. The three curves are for different ρ

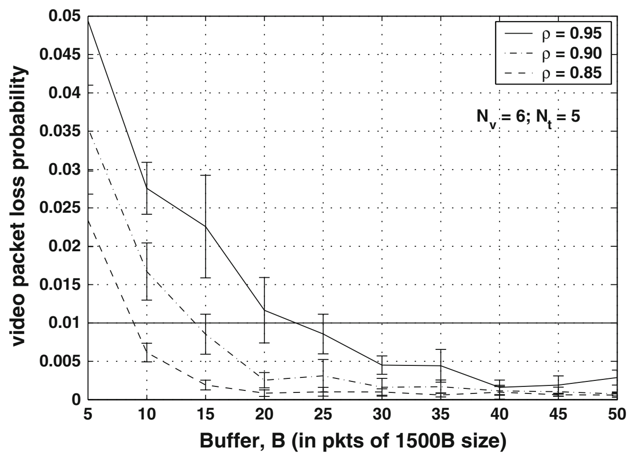

Fig. 12 Simulation results showing video packet loss probability vs. B, for $\varepsilon = 0.01$. $N_v = 6$, $N_t = 5$ and we have 4 non-synchronized CBR video sessions aggregating to the three different values of ρ

We now consider the situation in which video traffic comprises four non-synchronized CBR streams. Note that since the CBR streams are not synchronized, the net input at the video queue of the AP will be burstier than CBR. Figure 12 shows the plot of video packet loss probability vs. B for $\varepsilon = 0.01$, when $N_v = 6$ and $N_t = 5$, as obtained from the simulations, in such a case. We note that to ensure the loss probability below 1%, we need $B = 14$ for $\rho = 0.90$, which is less than that obtained with Poisson arrivals (i.e., $B = 28$).

We conclude that our analytical model provides a useful approach for sizing the buffer since it overestimates the required buffer by only a few packets. We find that a 50 packet buffer size, that translates to 75 KB, is more than sufficient for handling the video streaming sessions while guaranteeing the loss probability constraint (of less than 1%).

7 Conclusion

In this paper, we evaluated the performance of EDCA WLAN, when the traffic consists of VoIP calls, streaming video sessions and TCP download transfers. The analysis proceeds by modeling the evolution of the number of contending QSTAs at channel slot boundaries. This yields a Markov renewal process. A regenerative analysis then yields the required performance measures like the VoIP capacity, video saturation throughput and the TCP aggregate download throughput. The model predicts the measures that compare closely with the simulation results.

By an effective bandwidth approach we obtained the buffer size of QAP_{vd} that ensures the probability of loss of video packets to be within 1%.

Our work provides the following modeling insights:

- The idea of using saturation attempt probabilities as state dependent attempt rates yields an accurate model in the unsaturated case.
- Using this approximation, an IEEE 802.11e infrastructure WLAN can be well modeled by a multidimensional Markov renewal process embedded at channel slot boundaries.

We also obtain the following performance insights:

- Unlike the original DCF, the EDCA mechanism supports the coexistence of VoIP connections, video streams and TCP file transfers; but even one video streaming session and one TCP transfer reduces the VoIP capacity from 12 calls to 6 calls. Subsequently the VoIP capacity is independent of the number of video sessions and TCP transfers (see Figs. 4 and 5).
- For an 11 Mbps PHY, the net video throughput reduces linearly by 0.38 Mbps per additional VoIP call and when both VoIP and video sessions are present, the TCP file download throughput reduces linearly with the number of voice calls by 0.14 Mbps per additional VoIP call.
- By using a small buffer for AC 2 of AP (about 75 KB), the video packet loss probability can be kept within permissible limits (i.e., ≤ 0.01).

In related work, we have also provided an analytical model for IEEE 802.11e infrastructure WLANs, with voice being carried in contention period using HCCA, and TCP data in the remaining time using EDCA (see [29]).

Acknowledgment This work is based on research sponsored by Intel Technology, India.

Appendices

Appendix A: Expressions for various probability functions (defined in 3.2)

Define

$$\tau^{(\cdot)} := \beta^{(\cdot)}_{Y_j^{(v)}+1,1,Y_j^{(t)}+1}$$

Then,

$$\eta_v(Y_j^{(v)}, Y_j^{(t)}) = (1-\tau^{(v)})^{Y_j^{(v)}+1}$$

$$\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) = (1-\tau^{(vd)})$$

$$\eta_t(Y_j^{(v)}, Y_j^{(t)}) = (1-\tau^{(t)})^{Y_j^{(t)}+1}$$

$$\alpha_v(Y_j^{(v)}, Y_j^{(t)}) = Y_j^{(v)} \frac{\tau^{(v)}\eta_v(Y_j^{(v)}, Y_j^{(t)})}{1-\tau^{(v)}}$$

$$\alpha_t(Y_j^{(v)}, Y_j^{(t)}) = Y_j^{(t)} \frac{\tau^{(t)}\eta_t(Y_j^{(v)}, Y_j^{(t)})}{1-\tau^{(t)}}$$

$$\sigma_v(Y_j^{(v)}, Y_j^{(t)}) = \frac{\alpha_v(Y_j^{(v)}, Y_j^{(t)})}{Y_j^{(v)}}$$

$$\sigma_{vd}(Y_j^{(v)}, Y_j^{(t)}) = 1 - \eta_{vd}(Y_j^{(v)}, Y_j^{(t)})$$

$$\sigma_t(Y_j^{(v)}, Y_j^{(t)}) = \frac{\alpha_t(Y_j^{(v)}, Y_j^{(t)})}{Y_j^{(t)}}$$

$$\zeta_v(Y_j^{(v)}, Y_j^{(t)}) = \sum_{i=2}^{Y_j^{(v)}+1} \frac{\binom{Y_j^{(v)}+1}{i}(\tau^{(v)})^i \eta_v(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(v)})^i}$$

$$\zeta_t(Y_j^{(v)}, Y_j^{(t)}) = \sum_{i=2}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(v)})^i}$$

$$\psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)}) = \sum_{i=1}^{Y_j^{(v)}+1} \frac{\binom{Y_j^{(v)}+1}{i}(\tau^{(v)})^i \eta_v(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(v)})^i} \times \sum_{i=1}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})^i}$$

$$\psi_{v-vd}(Y_j^{(t)}, Y_j^{(t)}) = \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)}) \times \sum_{i=1}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})^i}$$

$$\begin{aligned}\psi_{vdAP}(Y_j^{(v)}, Y_j^{(t)}) = \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)}) \Big[& \eta_t(Y_j^{(v)}, Y_j^{(t)}) \\ & \times \sum_{i=1}^{Y_j^{(v)}+1} \frac{\binom{Y_j^{(v)}+1}{i}(\tau^{(v)})^i \eta_v(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(v)})^i} \\ & + \eta_v(Y_j^{(v)}, Y_j^{(t)}) \\ & \times \sum_{i=1}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})^i} \\ & + \psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)}) \Big]\end{aligned}$$

$$\begin{aligned}\psi_{tAP}(Y_j^{(v)}, Y_j^{(t)}) = \tau^{(t)} \Bigg[& \frac{\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})} \\ & \times \sum_{i=1}^{Y_j^{(v)}+1} \frac{\binom{Y_j^{(v)}+1}{i}(\tau^{(v)})^i \eta_v(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(v)})^i} \\ & + \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) \\ & \times \sum_{i=1}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})^i} \\ & + \psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)}) \\ & + \frac{\eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})}\psi_{v-vd}(Y_j^{(v)}, Y_j^{(t)}) \\ & + \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})\psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)}) \\ & + \sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})\eta_v(Y_j^{(v)}, Y_j^{(t)}) \\ & \times \sum_{i=1}^{Y_j^{(t)}} \frac{\binom{Y_j^{(t)}}{i}(\tau^{(t)})^i \eta_t(Y_j^{(v)}, Y_j^{(t)})}{(1-\tau^{(t)})^i} \Bigg]\end{aligned}$$

Note that all the probability functions are denoted as functions of $Y_j^{(v)}$ and $Y_j^{(t)}$ even when one of them may not be there in the expression, since β and hence τ is a function of both $Y_j^{(v)}$ and $Y_j^{(t)}$.

Appendix B: Numerical calculation of stationary distribution (refers to Sect. 3.2)

The transition probability matrix can be numerically generated using the above probability functions and distributions of arrivals of VoIP packets. For instance, consider $N_v = 5$, $N_t = 10$ and $N_{vd} = 1$. Let $(Y_j^{(v)}, Y_j^{(t)}, C_j) = (3, 2, 0)$ be the state of the Markov chain $\{Y_j^{(v)}, Y_j^{(t)}, C_j; j \geq 0\}$ at the end of jth channel slot. Then all three types of AC categories can contend in the next channel slot, implying that QAP_v, QAP_{vd}, QAP_t, 3 $QSTA_v$s and 2 $QSTA_t$s may contend for the channel in the $(j+1)$th channel slot.

Now let $C_{j+1} = 0$. This implies that an idle slot has occurred because none of the nodes contended for the channel. Then the number of contending $QSTA_t$s does not change. The number of contending $QSTA_v$s cannot decrease, but may increase by at most 2 (due to new arrival of packets). Then the state at $(j+1)$th channel slot boundary can be one of the 3 states : (3,2,0), if no VoIP packet arrives, (4,2,0), if one VoIP packet arrives, and (5,2,0), if 2 VoIP packets arrive. Then the transitional probabilities are as under:

$$Pr((3,2,0)|(3,2,0)) = \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})$$
$$\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})Pr\Big(B_{j+1}^{(v)} = 0|(Y_j^{(v)} = 3; L_{j+1} = \delta)\Big)$$

$$Pr((4,2,0)|(3,2,0)) = \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})$$
$$\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})Pr\Big(B_{j+1}^{(v)} = 1|(Y_j^{(v)} = 3; L_{j+1} = \delta)\Big)$$

$$Pr((5,2,0)|(3,2,0)) = \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})$$
$$\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})Pr\Big(B_{j+1}^{(v)} = 2|(Y_j^{(v)} = 3; L_{j+1} = \delta)\Big)$$

Instead, if $C_{j+1} = 1$, then this implies that an activity has occurred in the channel and that could have been either a successful transmission by one of the contending nodes or there has been collision between two or more contending nodes. Then the next states could be one of the these 10 states: (2,2,1) if $QSTA_v$ succeeded and no VoIP packet arrived; (3,2,1) if collision took place and no VoIP packet arrived or QAP_v succeeded and no VoIP packet arrived or QAP_{vd} succeeded and no VoIP packet arrived or $QSTA_v$ succeeded and 1 VoIP packet arrived; (4,2,1) if collision took place and 1 VoIP packet arrived or QAP_v succeeded and 1 VoIP packet arrived or QAP_{vd} succeeded and 1 VoIP packet arrived or $QSTA_v$ succeeded and 2 VoIP packets arrived; (5,2,1) if collision took place and 2 VoIP packets arrived or QAP_v succeeded and 2 VoIP packets arrived or QAP_{vd} succeeded and 2 VoIP packets arrived; (3,3,1) if QAP_t succeeded and no VoIP packet arrived; (4,3,1) if QAP_t succeeded and 1 VoIP packet arrived; (5,3,1) if QAP_t succeeded and 2 VoIP packets arrived; (3,1,1) if $QSTA_t$ succeeded and no VoIP packet arrived; (4,1,1) if $QSTA_t$ succeeded and 1 VoIP packet arrived; and (5,1,1) if $QSTA_t$ succeeded and 2 VoIP packets arrived. The transition probabilities for these transitions can similarly be written (as for $C_{j+1} = 0$ case) using the probability functions and conditional probability function of VoIP packet arrivals.

Thus the transition probability matrix can be numerically worked out and then, combining with $\sum_{n_v=0}^{N_v}\sum_{n_t=0}^{N_t}\sum_{c=0}^{1}\pi_{n_v,n_t,c} = 1$, the stationary distribution $\boldsymbol{\pi}$ of the Markov chain $\{Y_j^{(v)}, Y_j^{(t)}, C_j; j \geq 0\}$ can be evaluated.

Appendix C: Mean cycle length, L_j (refers to Sect. 3.3)

$$\mathsf{E}L_{j+1}|(C_j = 0)$$
$$= \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{s-v}\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\Big((\alpha_v(Y_j^{(v)}, Y_j^{(t)})$$
$$+ \sigma_v(Y_j^{(v)}, Y_j^{(t)})\Big)$$
$$+ T_{s-vdAP}\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_t(Y_j^{(v)}, Y_j^{(t)})\sigma_{vd}(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{s-tAP}\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\sigma_t(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{s-tSTA}\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\alpha_t(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{c-short}\eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\zeta_t(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{c-voice}\Big(\eta_t(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\zeta_v(Y_j^{(v)}, Y_j^{(t)})$$
$$+ \eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\psi_{v-tsta}(Y_j^{(v)}, Y_j^{(t)})\Big)$$
$$+ T_{c-vd}\psi_{vd-AP}(Y_j^{(v)}, Y_j^{(t)}) + T_{c-long}\psi_{tAP}(Y_j^{(v)}, Y_j^{(t)})$$

and

$$\mathsf{E}L_{j+1}|(C_j = 1)$$
$$= \eta_v(Y_j^{(v)}, Y_j^{(t)})\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{s-v}\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})(\alpha_v(Y_j^{(v)}, Y_j^{(t)}) + \sigma_v(Y_j^{(v)}, Y_j^{(t)}))$$
$$+ T_{c-voice}\eta_{vd}(Y_j^{(v)}, Y_j^{(t)})\zeta_v(Y_j^{(v)}, Y_j^{(t)})$$
$$+ T_{c-vd}\psi_{v-vd}(Y_j^{(v)}, Y_j^{(t)})$$

Note that the above Equations use L_j in units of system slots.

References

1. IEEE 802.11e Part 11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Amend-ment 8: Medium Access Control (MAC) Quality of Service Enhancements, 11 November 2005.
2. Choi, S., Prado, J., Nandgopalan, S. S. N., & Mangold, S. (2003). IEEE 802.11e contention-based channel access (EDCF) performance evaluation. *IEEE International Conference on Communications, 2*, 1151–1156.
3. Garg, P., Doshi, R., Greene, R., Baker, M., Malek, M., & Cheng, X. (2003). Using IEEE 802.11e MAC for QoS over wireless. In *Proceedings IEEE International Performance, Computing, and Communications Conference*, pp. 537–542, 9–11 April 2003.
4. He, D., & Shen, C. (2003). Simulation study of IEEE 802.11e EDCF. In *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual Volume 1*, pp. 685–689, 22–25 April 2003.
5. Banchs, A., Azcorra, A., Garcia, C., & Cuevas, R. (2005). Applications and challenges of the 802.11e EDCA mechanism: An experimental study. *IEEE Network, 19*(4), 52–58.
6. Blum, B., He, T., & Pointurier, Y. (2001). MAC layer abstraction for simulation scalability improvements. Dept. of Computer Science, University of Virginia, Charlottsville, VA, Tech. Rep., December 2001, available at http://www.cs.virginia.edu/bmb5v/cs851/report.pdf.
7. Ji, Z., Zhou, J., Takai, M., & Bagrodia, R. (2004). Scalable simulation of large-scale wireless networks with bounded inaccuracies. In *ACM MSWiM'04*, pp. 62–69, 4–6 October 2004.
8. Kroller, A., Pfisterer, D., Buschmann, C., Fekete, S. P., & Fischer, S. (2005). Shawn: A New Approach to Simulating Wireless Sensor Networks, http://arxiv.org/pdf/cs.DC/050200.
9. Ramaiyan, V., Kumar, A., & Altman, E. (2005). Fixed point analysis of single cell IEEE 802.11e WLANs: Uniqueness, multistability and throughput differentiation. In *Proceedings ACM Sigmetrics, '05. Journal version accepted for publication in IEEE TON, '08.*

10. Kuo, Y., Lu, C., Wu, E., & Chen, G. (2003). An admission control strategy for differentiated services in IEEE 802.11. In *Global Telecommunications Conference (GLOBECOM '03)*, Vol. 2, pp. 707–712, 1–5 December 2003.
11. Robinson, J., & Randhawa, T. (2004). Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function. *IEEE Journal on Selected Areas in Communications, 22*(5), 917–928.
12. Zhu, H., & Chlamtac, I. (2003). An analytical model for IEEE 802.11e EDCF differential services. In *Proceedings of The 12th International Conference on Computer Communications and Networks*, pp. 163–168, 20–22 October 2003.
13. Kong, Z., Tsang, D. H. K., Bensaou, B., & Gao, D. (2004). Performance analysis of IEEE 802.11e contention-based channel access. *IEEE Journal on Selected Areas in Communications, 22*(10), 2095–2106.
14. Shankar, S., del Prado Pavon, J., & Wienert, P. (2004). Optimal packing of VoIP calls in an IEEE 802.11 a/e WLAN in the presence of QoS constraints and channel errors. In *Global Telecommunications Conference, GLOBECOM '04*, Vol. 5, pp. 2974–2980, 29 November–3 December 2004.
15. Kumar, A., Altman, E., Miorandi, D., & Goyal, M. (2005). New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs. in *Proceedings of IEEE INFOCOM 2005*, pp. 1550–1561, 13 – 17 March 2005.
16. Clifford, P., Duffy, K., Foy, J., Leith, D. J., & Malone, D. (2006). Modeling 802.11e for Data traffic parameter design. In *WiOPT '06*, 3–7 April 2006.
17. Harsha, S., Kumar, A., & Sharma, V. (2006). An analytical model for the capacity estimation of combined VoIP and TCP file transfers over EDCA in an IEEE 802.11e WLAN. In *IEEE IWQoS '06*, pp. 178–187, 19–21 June 2006.
18. Wolff, R. W. (1989). *Stochastic modelling and the theory of queues*. Eaglewood Cliffs, NJ: Prentice Hall.
19. Kulkarni, V. G. (1995). *Modeling and analysis of stochastic systems*. London, UK: Chapman and Hall.
20. Kuriakose, G., Harsha, S., Kumar, A., & Sharma, V. (2008). Analytical models for capacity estimation of IEEE 802.11 WLANs using DCF for internet applications. In *Wireless Networks*. Springer Netherlands. doi:10.1007/s11276-007-0051-8.
21. Li, M., Claypool, M., & Kinicki, B. (2002). MediaPlayer versus RealPlayer—a comparison of network turbulence. In *Proc. ACM Sigcomm Internet Measurement Workshop*, pp. 131–136, November 2002.
22. Oelbaum, T., Baroncini, V., Tan, T. K., & Fenimore, C. (2004). *Subjective Quality Assessment of the Emerging AVC/H.264 Video Coding Standard*, http://www.itl.nist.gov/div895/papers/IBC-Paper-AVC%20VerifTestResults.pdf.
23. IEEE 802.11 WG, Reference number ISO/IEC 8802–11:1999(E) IEEEStd 802.11, 1999 edition.
24. Wiethlter, S., & Hoene, C. (2006). An IEEE 802.11e EDCF and CFB Simulation Model for ns-2. In http://www.tkn.tu-berlin.de/research/802.11e_ns2.
25. Prabhakaran, M. (2004). Design and Analysis of Link Scheduling Algorithms for Wireless Networks. Master's thesis, Indian Institue of Science, Bangalore.
26. Tickoo, O. & Sikdar, B. (2004). A queueing model for finite load IEEE 802.11 random access MAC. In *IEEE ICC, 2004*, Vol. 1, pp. 175–179, 20–24 June 2004.
27. Kumar, A., Manjunath, D., & Kuri, J. (2004). *Communication networking: An analytical approach*. San Francisco: Morgan-Kaufmann (an imprint of Elsevier), May 2004.
28. Dembo, A., & Zeitouni, O. (1998). *Large deviations techniques and applications* (2nd ed.). Springer.
29. Harsha, S., Anand, S., Kumar, A., & Sharma, V. (2006). An analytical model for capacity evaluation of VoIP on HCCA and TCP file transfers over EDCA in an IEEE 802.11e WLAN. In *International Conference on Distributed Computing and Networking (ICDCN '06)*, pp. 245–256, 27–30 December 2006.
30. Harsha, S. (2006). Master of Engg Thesis: Performance of Multimedia Applications on IEEE 802.11e Wireless LANs, ECE Department, Indian Institue of Science, Bangalore.

Author Biographies

Sri Harsha received his B.Sc. degree from Jawaharlal Nehru University (JNU), India, in 1994, B.Tech. degree in Telecommunications and Information Technology again from JNU in 2002 and an M.E. degree in Telecommunications from Indian Institute of Science (IISc), Bangalore, in 2006. His research interests include system-level analysis and design, and QoS provisioning in wireless networks.

Anurag Kumar (B.Tech., IIT Kanpur, Ph.D. Cornell University, both in EE) was with Bell Labs, Holmdel, for over 6 years. He is now a Professor in the ECE Department at the Indian Institute of Science (IISc), Bangalore, and also Chair of Electrical Sciences Division, IISc. His area of research is communication networking, and he has recently focused primarily on wireless networking. He is a Fellow of the IEEE, of the Indian National Science Academy (INSA), and of the Indian National Academy of Engineering (INAE). He was an associate editor of IEEE Transactions on Networking, and of IEEE Communications Surveys and Tutorials. He is a coauthor of the graduate text-books "Communication Networking: An Analytical Approach," and "Wireless Networking," both by Kumar, Manjunath and Kuri, and both published by Morgan-Kaufman/Elsevier.

Vinod Sharma completed B.Tech. in EE from IIT Delhi in 1978 and Ph.D. in ECE from Carnegie Mellon University at Pittsburgh in 1984. Since then he has worked in Northeastern University at Boston (1984–1985), University of California at Los Angeles (1985–1987) and Indian Institute of Science at Bangalore (1988–) where he is currently a Professor. Vinod Sharma's research interests are in Communication Networks and Wireless Communications.