Voice Capacity under Quality Constraints for IEEE 802.11a based WLANs

Niranjan Shetty
Department of Electrical and Computer Engineering
University of California, Santa Barbara, 93106.
niranjan@ece.ucsb.edu

Sayantan Choudhury
Department of Electrical and Computer Engineering
University of California, Santa Barbara, 93106.
sayantan@ece.ucsb.edu

Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara, 93106.
gibson@ece.ucsb.edu

ABSTRACT

The communication of voice over wireless local area networks (WLANs) is influenced by the choice of speech codec, packetization interval and PHY layer bit rates. These choices affect the number of voice users that can be supported on the WLAN as well as the speech quality experienced by each user. We investigate the effect of different combinations of these parameters for a 802.11a WLAN in different channel conditions with the objective of maximizing the number of voice users supported on the WLAN subject to a quality constraint. We use an indicator for assessing the speech quality experienced by a single user in a WLAN, based on a Perceptual Evaluation of Speech Quality (PESQ)-Mean Opinion Score (MOS) constraint and the probability of a voice user achieving this constraint. The contributions of this paper are three-fold. First, a PHY layer rate adaptation scheme is proposed, in which the operating rate for each Signal to Noise Ratio (SNR) is chosen as the one that maximizes the capacity given a quality constraint. Second, based on the PHY layer rate adaptation scheme, we evaluate the effect of the choice of codec and voice payload size on the capacity values obtainable at different SNRs. Finally, we show the effect of channel conditions and the tightening of quality constraints on the capacity values at different SNRs.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: [wireless communication]

General Terms

Design, Performance

Keywords

Voice over WLANs, voice quality indicator, rate adaptation, number of voice users, frequency selective multipath fading, PESQ MOS, IEEE 802.11a.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IWCMC'06, July 3–6, 2006, Vancouver, British Columbia, Canada.
Copyright 2006 ACM 1-59593-306-9/06/0007 ...$5.00.

1. INTRODUCTION

The widespread use of Voice over IP (VoIP) and its extension to wireless local area networks (WLANs) has led to an increased interest in the study of voice over wireless LANs (VoWLANs). Since conventional WLANs have been designed for packet data, communicating voice over WLANs has its own challenges. Packet voice communication is sensitive to delay, but relatively less sensitive to packet losses compared to communication of data packets. In the case of VoWLANs, multiple choices for speech codec, the packetization interval and the PHY layer bit rate are available. These factors affect both the quality that a voice user in the network can expect to experience, and the number of voice users that can be supported on an access point, also defined as its capacity [5].

Faced with multiple choices, the issue is one of finding the best combination of voice payload size, transmitted bit rate and speech codec that achieves the maximum capacity with good quality under realistic channel conditions. In this paper, we investigate such combinations for an 802.11a based WLAN with a PESQ-MOS based quality constraint for frequency selective multipath fading channels. An extensive set of experiments for different multipath fading realizations were conducted for different combinations of codecs (G.711 and G.729), voice payload sizes (10 ms and 20 ms with and without header compression), data rates (6 Mbps-54 Mbps) and SNRs (0 dB to 40 dB in increments of 5 dB). Based on an analysis of this experimental data, we

- Compare the PESQ-MOS performance of each codec and payload size for different SNRs for every rate offered by 802.11a
- Employ a quality constraint based on obtaining good speech quality (e.g. MOS ≥ 3.0) with a very high probability (e.g > 0.9) that accounts for the variation in MOS with packet loss patterns
- Devise a PHY layer bit rate adaptation scheme which determines the set of operating rates for different SNRs based on maximizing the speech quality constrained capacity
- Compare the performance of G.711 and G.729 using voice payload sizes of 10 ms and 20 ms based on the

capacity values obtained using the above rate adaptation scheme for different SNRs.

- Contrast capacity in a multipath fading environment with that obtained in an AWGN channel.
- Evaluate the effect on capacity of tightening the quality constraints.

Previous work on the capacity of voice based wireless LANs evaluate capacity by changing the voice payload size and codec under a fixed operating rate [5] and does not examine the variation in speech quality with different packet loss patterns corresponding to the same packet loss rate (PLR) [12, 13]. In [5], it is shown that the capacity for an IEEE 802.11b network at a data rate of 11 Mbps can be increased with an increase in the allowable delay. However, the quality metric used in [5] is based on an average MOS score which, as we demonstrate, is a poor reflection of the quality an individual user can expect to achieve. Furthermore, in [5] the MOS values for a packet size of 50 ms for different packet PLRs is assumed to be equal to those for 20 ms. This ignores the fact that large voice payload sizes are more difficult to conceal and hence result in poorer performance compared to smaller voice payloads at the same PLRs.

The effect of fading in 802.11a based voice networks is considered in [14]. Their use of a MOS constraint and the probability of not achieving this constraint, as an indicator of quality, is similar to the one employed by us. However, the MOS values used in [14] are based on an average PLR and are obtained from the E-Model, and do not incorporate the effect of variation in MOS with packet loss patterns as we do here. They also do not examine the capacity for different rates or different packet sizes.

In [12], the capacity of the network was evaluated for different rates, packetization intervals, and different 802.11 standards. However their analysis considers the G.711 codec only and is based on a PLR constraint of $\leq 2\%$. In [13], the authors use results from [3] to estimate the PLR vs SNR characteristics. However, these PLRs were estimated for 512 byte payload sizes, which are much longer than 10 and 20 ms G.711 and G.729 payload sizes.

The work in this paper is a comprehensive effort that includes defining a quality constraint for WLAN capacity evaluation, developing a PHY layer rate adaptation scheme based on quality and capacity, and evaluating the quality-constrained capacity for different payload sizes, codecs, transmitted bit rates and different SNRs. Each of these are discussed in detail in the following sections.

2. SPEECH QUALITY INDICATOR FOR VOICE OVER WIRELESS LANS

The popular speech quality metrics currently in use for narrowband speech are the E-model [8] and the PESQ [11]. PESQ uses the reference and degraded speech to provide an MOS score for the degraded speech based on a perceptual evaluation. It does not incorporate the effect of delay but reflects the effect of packet loss patterns and the packet loss concealment scheme on the overall quality. In Figure 1 we evaluate the PESQ-MOS for two speech files of male and female speech, respectively, each of which is 8 seconds long and is coded using G.729 with a voice payload size of 10 ms. 500 different packet loss patterns are considered for each of a set of PLRs from from 1% to 10 %. We observe that for a specific PLR, there can be a significant variation in the PESQ-MOS scores. For example, at a PLR of 10 %, the PESQ-MOS lies between 2.7 and 3.3. This is because the perceptual importance is different for different frames in the speech file. For different packet loss patterns corresponding to a given PLR, the PESQ-MOS varies depending on the perceptual importance of the frames lost.

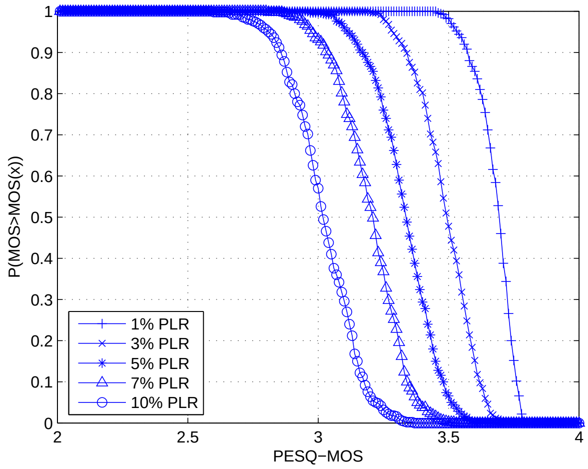

Figure 1: CCDF of PESQ MOS values for G.729 coded speech with frame size of 10 ms for different PLR based on 500 different packet loss patterns corresponding to each PLR

Based on the above complementary cumulative distribution functions (CCDFs), we can define a quality indicator called the MOS_x as the MOS value a user can expect to obtain or exceed with a probability of x%. Based on this definition, MOS_{50} refers to the value of MOS which a user can expect to obtain or exceed with a probability of 0.5. The values of the average MOS, MOS_{50} and MOS_{90} over the 500 realizations corresponding to each PLR, are plotted in Figure 2. We observe that the average value of the MOS is approximately equal to MOS_{50}. This implies that the average MOS value represents a MOS score that a user can expect to obtain with a probability of only 0.5. In other words, the average MOS value does not guarantee that a user will actually experience that quality. Alternately we can specify a quality indicator with a higher guarantee, such as the MOS_{90} which is defined as the MOS value that a user can expect to exceed with a probability of 90%. This indicator provides a value of the MOS that a user can expect to obtain or exceed with a high probability, unlike the average MOS.

The E-model evaluates the speech quality based on a number of network and speech parameters, such as the loudness rating, codec used, PLR and the delay on the network. It has the advantage of providing a real-time estimate of the speech quality and also incorporates the effect of delay. However the E-model provides an estimate of quality based on PLR. As we have shown, MOS values may vary depending on the packet loss patterns for a given PLR. This is not accounted for in the E-model.

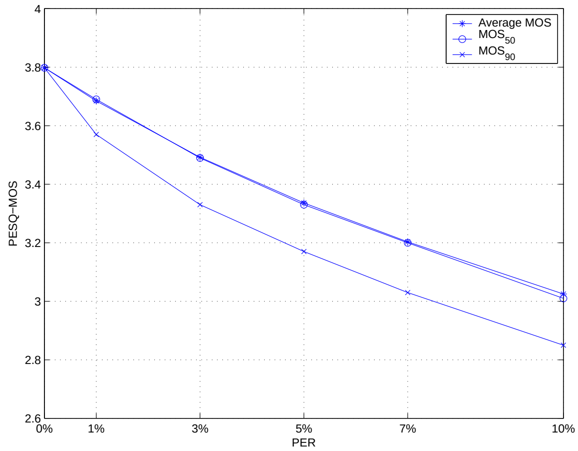

Figure 2: Comparison of average MOS, MOS_{50} and MOS_{90} for G.729 coded speech

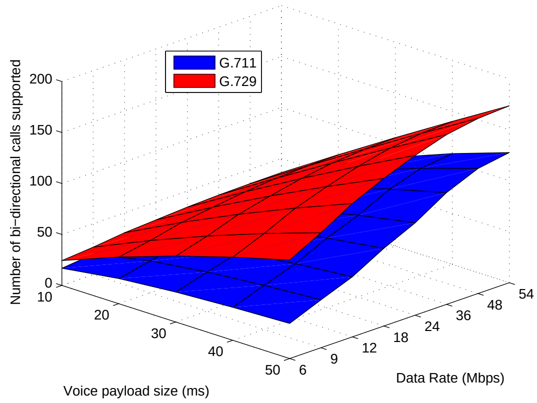

Figure 3: Analytical capacity bounds for 802.11a with RoHC for G.711 and G.729 coded speech for different voice payload sizes and data rates

Based on the above observations, we use the PESQ-MOS with a probability constraint as an indicator of quality in our experiments. It is important to note that the E-Model standard [8] includes mapping functions for mapping MOS values to E-Model values and vice versa. The use of the E-model might be appropriate if the effect of delay has to be incorporated. Since we consider small voice payload sizes of up to 20 ms and assume no retransmissions in our experiments, the delay would typically not exceed the acceptable threshold of 150 ms for allowable one-way delay [9].

3. SPEECH QUALITY CONSTRAINED CAPACITY IN FREQUENCY SELECTIVE MULTIPATH FADING

3.1 Analytical Expression of Capacity

The analytical expression for capacity is the same as used in [5] and is given by

$$N = \frac{T_{vp}}{T_{transmit}} \tag{1}$$

where T_{vp} is the size in ms of the voice payload, and $T_{transmit}$ is the time involved in the successful transmission of a packet in an 802.11a network,

$$T_{transmit} = 2 * (T_{voice} + SIFS + T_{ack} + DIFS) + (T_{slot} * CW_{min}/2)$$

T_{voice} is the time taken to transmit a packet at a data rate R Mbps. The SIFS, DIFS and T_{slot} times for 802.11a are 16 μs, 34 μs and 9 μs, respectively, while the minimum contention window size (CW_{min}) is 15 [7]. T_{ack} is the time that the sender has to wait to receive an acknowledgment from the receiver and depends on the data rate R. Thus the capacity value varies with data rates supported by 802.11a and the voice payload size in each packet. A plot of the capacity values for different rates and voice payload sizes for G.711 and G.729 is shown in Figure 3.

It is important to note that the above definition of capacity assumes ideal channel conditions and no retransmissions. For such an ideal case, it is seen from Figure 3 that maximum analytical capacity is obtained by employing the G.729 speech codec and using the maximum available voice payload size and transmission rate. However, under realistic channel conditions, the PLR would depend on the transmission rates, payload sizes and the SNR. The quality of the speech is dependent both on the packet loss pattern and the codec employed. For a given SNR, the choice of data rate, codec, and voice payload length impacts both the capacity supported and the speech quality obtained by each user. Therefore, the objective of finding the combination of transmission rate, codec and voice payload that achieves the highest capacity given a quality constraint for different SNRs assumes significance. In the remainder of this paper, we define capacity to mean the number of users supported on an access point under a quality constraint.

3.2 Simulation scenario

The scenario we consider consists of multiple users associated with an access point which in turn is connected to a wired IP network. Each user associated with the access point is talking to another user through the IP network. Thus the number of users associated with the access point is equal to the number of bi-directional calls and is referred to as the capacity of the access point.

Two speech files each of duration 8 seconds and containing speech segments from a male and female speaker were used. For coding the speech, G.711 [1] with ITU packet loss concealment (PLC) [10] and G.729 with default PLC [2] have been considered. Voice payload sizes of 10 ms and 20 ms are used, which correspond to 800 packets and 400 packets, respectively. The use of longer voice payload size adds to the overall delay. We further observe from Figure 4 that a voice payload size of 50 ms provides a MOS of $<$ 3.0 at a PLR of slightly greater than 3%. This is due to the fact that larger voice payload sizes are more difficult to conceal by the speech decoder as compared to smaller voice payload sizes. Hence for our experiments, we do not consider voice payload sizes greater than 20 ms. Packetization of speech with and without Robust Header Compression (RoHC) has been considered. In case of the latter, the RTP/UDP/IP header size is reduced from 40 bytes to 2 bytes. No silence suppression is assumed.

In order to estimate the packet error rate under different

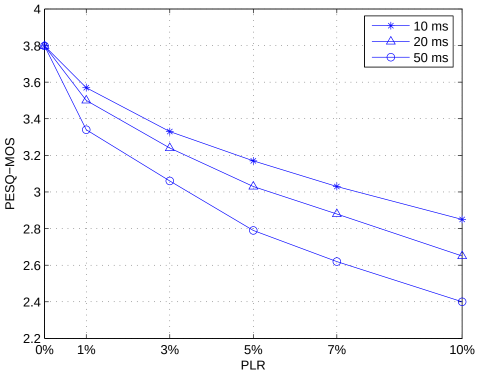

Figure 4: MOS_{90} **values for G.729 coded speech with voice payload sizes of 10 ms, 20 ms and 50 ms**

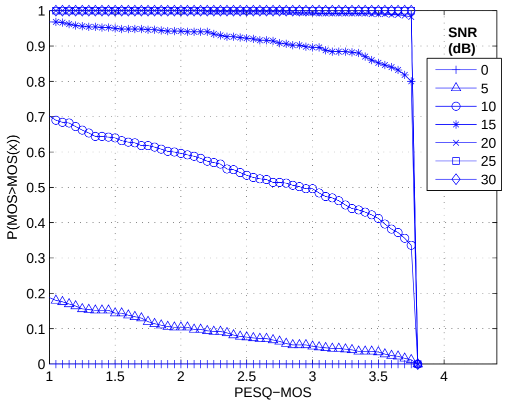

Figure 5: CCDFs of PESQ-MOS values for voice payload size of 20 ms coded using G.729 and sent on an 802.11a network at 24 Mbps with RoHC

channel conditions, we modified a readily available OFDM simulator for the IEEE 802.11a PHY [6]. Non-fading channels as well as multipath fading channels are considered. Noise is modeled as AWGN in both scenarios. The decoding at the receiver is based on soft decision Viterbi decoding. We also assume perfect synchronization and channel estimation.

The wireless channel model used for the multipath fading case is the Nafteli Chayat model [4], which is a standard indoor wireless channel model with an exponentially decaying Rayleigh faded path delay profile. The RMS delay spread used is 50 nanoseconds, which is typical for home and office environments. Each realization of the multipath delay profile corresponds to a stationary VoIP user which results in a specific loss pattern. 500 different multipath fading realizations are conducted for each data rate, average SNR, codec and packet size.

All the data rates for 802.11a were used and no retransmission is assumed. Values of SNR ranging from 0 dB to 40 dB in increments of 5 dB are employed. The packet loss patterns for each fading realization are used for dropping packets in the coded speech files with the specified packetization interval, and the PESQ-MOS is evaluated for the decoded speech.

3.3 Speech Quality Constraint

The MOS values for the 500 realizations for different SNRs and for different PHY layer rates, codecs and packet sizes, are plotted as CCDFs. The plot in Figure 5 for G.729 coded speech with voice payload size of 20 ms and a rate of 24 Mbps is one example from the set of plots obtained for different rates.

Based on the assumption that a MOS $\geq$ 3.0 achieves a good speech quality, we observe from the CCDF plots that corresponding to each rate, there is a set of low SNR values which satisfy this constraint with a probability of $\leq$ 0.1. At these SNRs, a user can expect good quality with a very low probability. In Figure 5, this is represented by the SNRs up to 5 dB. Similarly there are a set of high SNR values which can achieve the constraint with a probability of $\geq$ 0.9. At these SNRs, a user can expect to achieve good quality with a very high probability. In Figure 5, SNRs of 15 dB and higher can be classified in this category. In between these two regions, there exists a set of intermediate SNR values with a probability of between 0.1 to 0.9 of achieving the MOS constraint. This is illustrated by the SNR of 10 dB in Figure 5. As a means of ensuring that each user achieves good voice quality with a very high probability, a speech quality constraint of MOS $\geq$ 3.0 with a probability of $\geq$ 0.9 is employed. This defines our region of interest as being those SNRs for each rate which satisfy this constraint.

3.4 Rate Adaptation Scheme based on maximization of Quality-constrained Capacity

Based on the constraint of obtaining a MOS $\geq$ 3.0, we obtain the probabilities of achieving this constraint for different SNRs corresponding to each PHY layer rate, and for different such rates for both the G.711 and G.729 speech codecs. An example plot of the probabilities for different PHY rates and with varying SNR is shown in Figure 6 for G.729 coded speech with a voice payload size of 20 ms and using RoHC. A PHY layer rate adaptation scheme is devised, that for each SNR we choose the rate, which satisfies the speech quality constraints, and provides the highest capacity. In Figure 6, the points where the horizontal line corresponding to a probability of 0.9 intersects the curves for each rate represents the set of operating rates for the corresponding SNRs on the x-axis. The only exception is the rate of 9 Mbps, which exhibits a performance worse than 12 Mbps and hence is never employed at any SNR. The approximate values of the SNRs at which different 802.11a rates are employed under the proposed rate adaptation scheme for G.729 with a voice payload size of 20 ms, and the corresponding capacity values are shown in Table 1.

Table 1: Rate adaptation scheme for 802.11a using G.729 with a payload size of 20 ms and RoHC under multipath fading

Rate (Mbps)	6	12	18	24	36	48	54
SNR (dB)	7-9	9-14	14-15	15-20	20-23	23-24	>24
Capacity	45	58	62	66	68	71	71

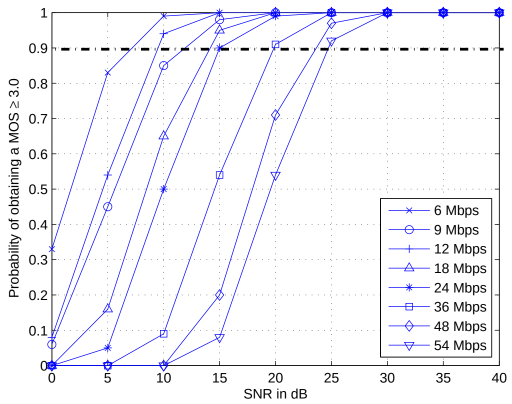

Figure 6: Probability of users experiencing a MOS > 3.0 for G.729 coded speech with voice payload size of 20 ms at different rates and SNRs

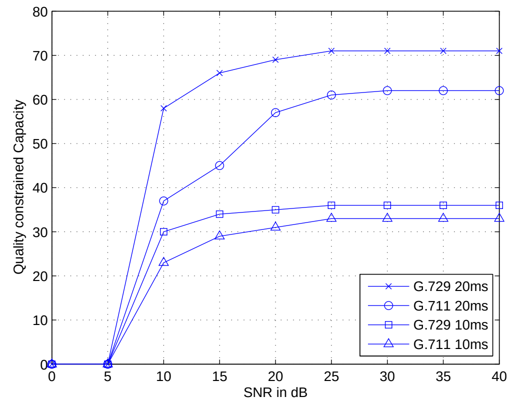

Figure 7: Capacity for G.729 and G.711 for voice payload sizes of 10ms and 20ms for varying SNR values subject to the quality constraint of achieving a MOS ≥ 3.0 with a probability ≥ 0.9, for frequency selective multipath fading

3.5 Capacity values for different rates, codecs and frame sizes at different SNRs

Using the rate adaptation scheme discussed above, we can associate a data rate with each SNR under the specified quality constraints. For a given codec and voice payload size, a capacity value is associated with each data rate as shown in Figure. 7. This allows us to associate a capacity value with each SNR. The capacity under the specified quality constraints for G.711 and G.729 for voice payload sizes of 10 ms and 20 ms using RoHC are plotted in Figure 7 for a range of SNR values. We observe that G.729 with a voice payload size of 20 ms provides the highest capacity followed by G.711 at 20 ms, G.729 at 10ms and G.711 at 10ms. The use of 20 ms voice payload sizes results in almost a doubling of the capacity values at SNRs $\geq$ 25 dB. We also observe that the capacity values for G.711 and G.729 are close for a voice payload size of 10 ms with G.729 having a slightly higher capacity. This gap widens significantly when the voice payload size is increased to 20 ms.

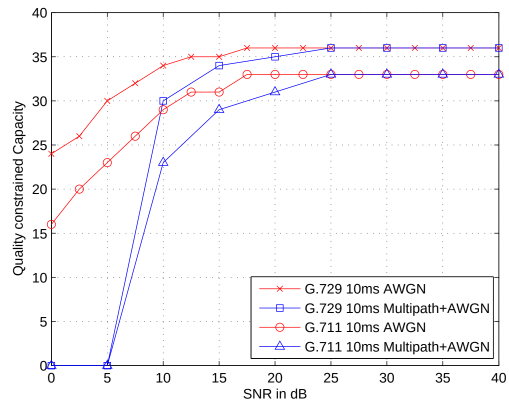

Figure 8: Capacity for G.729 and G.711 for voice payload size of 10 ms with and without frequency selective multipath fading subject to a speech quality constraint of achieving a MOS ≥ 3.0 with a probability of ≥ 0.9

4. EFFECT OF CHANNEL MODEL AND QUALITY CONSTRAINTS ON CAPACITY

4.1 Performance comparison with an AWGN channel

The effect of the channel model can be evaluated by considering an AWGN channel for G.711 and G.729 with RoHC and a voice payload size of 10 ms. From an analysis of the CCDF curves for different SNRs for each rate of 802.11a (not shown), we find that the variance in AWGN-only for the intermediate SNRs for each rate is relatively small as compared to multipath fading. Similar to the approach used for multipath fading, we devise a rate adaptation scheme based on achieving the maximum capacity given the quality constraints. This gives us a set of operating rates for different SNRs and corresponding capacity values. These quality-constrained capacity values are plotted in Figure 8 for a range of SNRs in an AWGN channel for speech coded using G.711 and G.729 using a packet size of 10 ms.

It is observed from Figure 8, that for an AWGN channel with an SNR of $\leq$ 20 dB, higher values of the capacity are obtained in comparison with multipath fading. At SNRs below 10 dB, while no users can be supported in the case of multipath fading, under conditions of no fading as many as 16 users with G.711 and 24 users with G.729 can be supported at an SNR of 0 dB. Thus, the type of channel model has a significant effect on the capacity values obtained at these SNRs. At SNR values $\geq$ 25 dB, the reliability of the channel is high and hence the capacity values for AWGN with and without multipath fading converge.

4.2 Performance comparison under stricter quality constraints

We previously chose our constraints as MOS $\geq$ 3.0 with a probability of $\geq$ 0.9. We now consider the effect of tightening the MOS constraints on the capacity. A tighter MOS constraint of achieving a MOS $\geq$ 3.5 is chosen, with a probability of 0.98 or greater of achieving this MOS. Under these

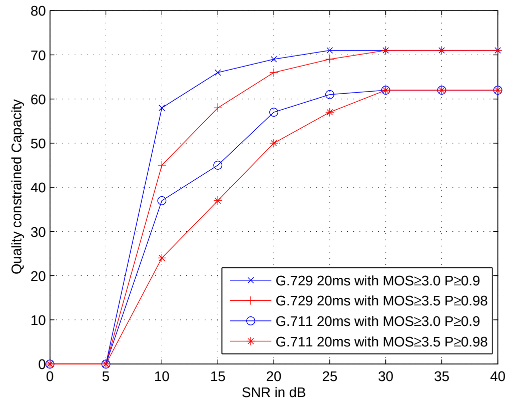

Figure 9: Capacity for G.729 and G.711 coded speech with voice payload size of 20 ms over frequency selective multipath fading channels subject to different speech quality constraints

new constraints, we obtain a different set of operating rates associated with the SNR values for each codec and voice payload size under multipath fading. The capacities for different SNRs under this new constraint are plotted alongside those for the original constraint in Figure 9. The difference in the calls supported is significant at an SNR of 10 dB with as many as 13 fewer users supported under the tighter quality constraints for G.729 coded speech with 20 ms voice payload size. With increasing SNR, the advantage of relatively looser quality constraints in achieving a higher capacity starts decreasing and at an SNR of 25 dB, the capacity values converge. Thus tightening the MOS constraints results in a reduction of capacity. However, each of the users supported under tighter quality constraints can expect a significantly better quality with a much higher probability.

We also observe that the capacity advantage obtained by the use of G.729 as compared to G.711 and the use of a voice payload size of 20 ms compared to 10 ms holds even under the tighter constraints of achieving a MOS ≥ 3.5 with a probability ≥ 0.98. It is important to note here that if the MOS constraints are tightened further to be ≥ 4.0, then G.711 is left as the only option since the constraint is greater than the PESQ-MOS of 3.8 of the G.729 coded speech files in the absence of packet losses.

5. CONCLUSIONS

Using a quality indicator that specifies the probability of achieving a chosen MOS value, we propose a PHY layer rate adaptation scheme that involves choosing a transmission rate that maximizes capacity while meeting the quality constraints. These quality-constrained capacity values for different SNRs are then used to compare different codecs and different voice payload sizes. We show that G.729 offers an advantage over G.711 in terms of the obtainable capacity under a quality constraint of obtaining a MOS ≥ 3.0 with a probability ≥ 0.9. Under the same speech quality constraint, the use of a voice payload size of 20 ms is seen to support a higher capacity as compared to a voice payload size of 10 ms. We also show that the capacity values depend on the channel model considered up to an SNR of 25 dB. Further tightening the quality constraints results in a reduced capacity up to an SNR of 25 dB, suggesting a tradeoff between providing better quality with higher guarantee to the user, and increasing the capacity associated with an access point.

6. ACKNOWLEDGMENTS

This work was supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc. and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc.

7. REFERENCES

[1] ITU-T Recommendation G.711. *Pulse Code Modulation for voice frequencies*, Nov. 1988.

[2] ITU-T Recommendation G.729. *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)*, Mar. 1996.

[3] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull and J. McGeehan. A Comparison of the HIPERLAN/2 and IEEE 802.11a Wireless LAN Standards. *IEEE Communications Magazine*, May 2002.

[4] N. Chayat. Tentative Criteria for Comparison of Modulation Methods. *IEEE P802.11-97/96*, Sept. 1997.

[5] D. P. Hole and F. A. Tobagi. Capacity of an IEEE 802.11b wireless LAN supporting VoIP. *IEEE Intl. Conf. on Communications*, 2004.

[6] J. Heiskala and J. Terry. *OFDM Wireless LANs: A Theoretical and Practical Guide* . Sams, December 2001.

[7] IEEE 802.11a WG. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification: High-Speed Physical Layer in the 5 GHz Band. 1999.

[8] ITU-T Recommendation G.107. The E-model: A computational model for use in transmission planning. Feb. 2001.

[9] ITU-T Recommendation G.114. One-Way Transmission Time. May 2000.

[10] ITU-T Recommendation G.711 Appendix I. A high quality low-complexity algorithm for packet loss concealment with g.711. Oct. 1999.

[11] ITU-T Recommendation P.862. PESQ: An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. Feb. 2001.

[12] K. Madepalli and P. Gopalakrishnan and D. Famolari and T. Kodama . Voice Capacity of IEEE 802.11b, 802.11a and 802.11g Wireless LANs. *GLOBECOM*, 2004.

[13] K. Madepalli and P. Gopalakrishnan and D. Famolari and T. Kodama . Voice Capacity of IEEE 802.11b and 802.11a Wireless LANs in the Presence of Channel Errors and Different User Data Rates. *VTC*, 2004.

[14] O. Awoniyi and F. A. Tobagi. Effect of Fading on the performance of VoIP in IEEE 802.11a. *IEEE Intl. Conf. on Communications*, 2004.