MOS_x and Voice Outage Rate in Wireless Communications

Sayantan Choudhury, Niranjan Shetty, and Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara
Email: {sayantan, niranjan, gibson}@ece.ucsb.edu

***Abstract*— Average packet loss rate (PLR) or average mean opinion score (MOS) are often used performance indicators for voice communications. However, even for a fixed average PLR, the delivered voice quality depends on the location of the packet losses as well as the distribution of packet losses due to the differences in voice codecs, their packet loss concealment schemes, the difficulty of concealing packet losses in unvoiced-to-voiced and voiced-to-unvoiced transition regions, and the difficulty of concealing successive packet losses. The result is that there is a distribution of achieved MOS values for a fixed average PLR and that the average MOS value may not capture the voice communications performance. Using the PESQ-MOS, we study the distribution of MOS values for wireless voice communications over additive white Gaussian noise (AWGN) and multipath fading channels using G.711 and G.729 voice codecs and their packet loss concealment schemes. Based on the distribution of PESQ-MOS, we define a quality indicator referred to as the MOS_x, which is the MOS value that a user can expect to achieve or exceed x% of the time, where MOS and x are chosen to correspond to an acceptable voice outage rate. The MOS_x is then used to evaluate the performance of G.711 and G.729 codecs over frequency selective fading and AWGN channels for voice over IEEE 802.11a wireless LANs.**

I. Introduction

In recent years, there has been significant interest in voice communications over wireless local area networks (VoWLANs). However, conventional WLANs have been designed for data traffic, and the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) protocol as well as the multipath fading channels have a significant effect on performance. Most voice quality assessment techniques focus on average packet loss rate (PLR) or average MOS and fail to consider the distribution of packet losses on delivered voice quality.

The effect of multipath fading on voice communications in 802.11a based networks was considered in [1], where it was shown that for certain signal to noise ratios (SNRs), the variation in the packet loss rate for a multipath fading channel was significant and that the average PLR was a poor indicator of the quality. The speech quality metric was based on an R-factor value obtained using the E-model and the PLR for different fading realizations, which was then mapped onto an MOS score.

In this paper, we investigate voice communications over IEEE 802.11a WLANs using PESQ-MOS, different voice codecs, and packet loss concealment (PLC) with different payload sizes and different data rates. We show that even for the same PLR, the speech quality varies depending on the pattern of packet losses. Hence, we define a speech quality indicator MOS_x, which is the MOS value that a user can expect to achieve or exceed x% of the time, and where MOS and x are chosen to correspond to an acceptable voice outage rate.

We perform a comprehensive voice quality assessment in frequency selective multipath fading and a comparison of G.711 and G.729 voice codecs for different payload sizes and supported data rates of IEEE 802.11a. We compare the voice quality and the SNR regions for the lowest and highest rate supported in IEEE 802.11a corresponding to 6 Mbps and 54 Mbps, respectively. We then compare and contrast the performance of G.711 with a higher intrinsic MOS and G.729 which allows a higher number of voice users [2]. We also contrast the performance of AWGN and multipath fading for the 6 Mbps data rate.

The paper is outlined as follows. In the next section, we provide a brief description of the frequency selective multipath fading channel model. Section III describes the packetization and voice codecs used. In Section IV, we introduce two concepts, namely, MOS_x and voice outage rate, and discuss their suitability as performance indicators for voice quality assessment in frequency selective multipath fading. In Section V, we conduct a performance evaluation of voice quality in AWGN and multipath fading channels for different combinations of codec (G.711 and G.729), payload sizes (10 ms and 20 ms) and different data rates (6 to 54 Mbps) supported by IEEE 802.11a WLANs. In Section VI, we compare the SNR thresholds for different link adaptation schemes. Section VII provides some conclusions.

II. Frequency Selective Multipath Fading Channel Model

The IEEE 802.11a PHY provides 8 modes with varying data rates from 6 to 54 Mbps by using different modulation and coding schemes. Forward error correction (FEC) is

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

done by using rate 1/2 convolutional coding and bit interleaving for the mandatory rates and using puncturing for the higher rates. A detailed description of OFDM systems and its applications to wireless LANs can be found in [3], [4].

In order to estimate the packet error rate under different channel conditions, we modified a readily available OFDM simulator for the IEEE 802.11a PHY [4]. The scenario we consider is of a single-user communicating with an access point or with another node in an ad-hoc or mesh network. 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 perfect channel estimation.

The wireless channel model used for the multipath fading case is the Nafteli Chayat model [5], which is a standard indoor wireless channel model with an exponentially decaying Rayleigh faded path delay profile. The impulse response of the channel, illustrated in Fig. 1, is composed of complex samples with uniformly distributed phase and Rayleigh distributed magnitude with an exponentially decaying average power profile. This is given by:

$$h_k = N(0, 1/2\sigma_k^2) + jN(0, 1/2\sigma_k^2) \quad (1)$$

$$\sigma_k^2 = \sigma_0^2 exp(-k * T_s/T_{rms}) \quad (2)$$

where $N(0, 1/2\sigma_k^2)$ is a zero mean Gaussian random variable with variance $1/2\sigma_k^2$ and σ_0^2 is chosen so as to ensure $\sum_k \sigma_k^2 = 1$ i.e., the channel coefficients are normalized so as to ensure the same average power. The rms delay spread used is 50 nanoseconds, which is typical for home and office environments.

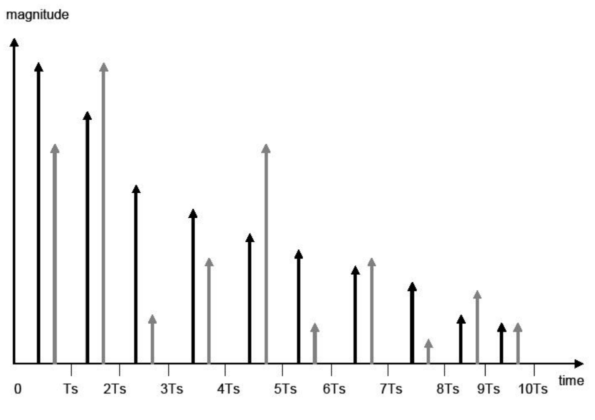

Fig. 1. Multipath channel impulse response showing average response (black) and an individual realization (grey)

III. PACKETIZATION AND VOICE CODECS

The G.711 PCM codec [6] with ITU Packet Loss Concealment (PLC) [7] and the G.729 codec with default PLC [8] are used in our simulations. The payload lengths used for our simulation purposes are 80 bytes and 160 bytes for 10 ms and 20 ms G.711 speech frames respectively, and 10 bytes and 20 bytes for 10 ms and 20 ms G.729 speech frames, respectively. The packet error distribution for each realization was obtained by transmitting 800 packets corresponding to 8 seconds of speech, comprised of one male and one female speech sentence. This was repeated for 500 random realizations for each set of average SNRs, PHY rate, voice payload size and fading channel. In addition to the voice payload, each packet contains the RTP/UDP/IP headers. These headers have a default size of 40 bytes. In our simulations, Robust Header Compression (RoHC) has been assumed. This reduces the RTP/UDP/IP header size from 40 bytes to 2 bytes. No silence suppression is used.

IV. MOS_x AND VOICE OUTAGE RATE

The Perceptual Evaluation of Speech Quality (PESQ) [9] is a popular objective measure for evaluating narrowband speech. Based on a perceptual comparison of the original and reference speech inputs, PESQ provides an objective value of the MOS. For a specified packet loss rate (PLR), the ***average*** value of the PESQ-MOS is typically used as the speech quality indicator. However, based on our experiments, we observe that this does not accurately reflect the quality that is experienced by a user. In Fig. 2 we evaluate the PESQ-MOS for two speech files of male and female speech, each 8 seconds long. 500 different packet loss patterns are considered for each of a set of packet loss rates (PLRs) from 1% to 10%. We observe that for a specific PLR, there is a significant variation in the PESQ-MOS scores due to variation in the perceptual importance of the lost packets.

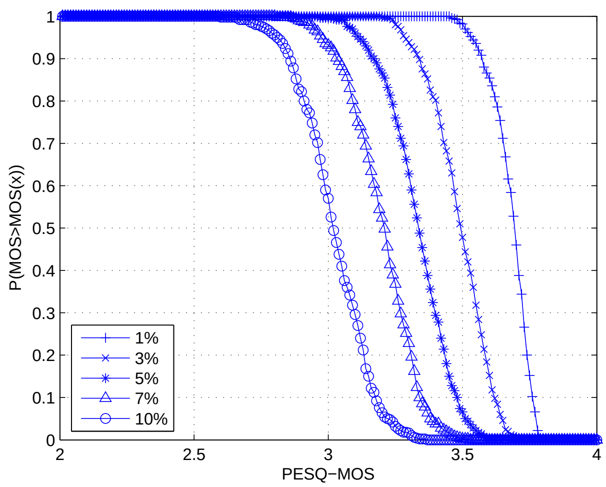

Fig. 2. CCDF of PESQ MOS values for G.729 coded speech with frame size of 10ms for different PLR

We introduce another measure of quality called the MOS_x, which is defined as the MOS value a user can expect to achieve for at least x% of the realizations e.g. MOS_{50} refers to the value of MOS which a user can expect to exceed for 50% of the realizations. The values of the average MOS, MOS_{50} and MOS_{90} are plotted in Fig. 3. 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 only 50% of the time. Thus, the average MOS value does not guarantee that a user will actually experience that quality. Alternately we define the MOS_{90} as the MOS value that a user can expect to exceed with a probability of 0.9. Based on this, we can define good quality speech as one in which the user achieves a high value of MOS for at least x% of the realizations (MOS_x). We consider a value of x equal to 90 as being a good guarantee for speech quality, and a MOS of 3.0 as reflecting good quality speech in our experiments.

The MOS specified by the E-model for 5% PLR using 10 ms G.729 is 3.3, which is just below the average MOS value plotted in Fig. 3. Hence, schemes mapping PLR to MOS using the E-model may not be a reliable estimator of the quality perceived by an user in a wireless network. In the E-model, the degradation caused by packet losses is accounted for in the effective equipment impairment factor (I_{e-eff}) [10]. This is comprised of the equipment impairment factor (I_e) that depends on the codec used, the packet loss probability (Ppl) and the packet loss robustness factor (Bpl). Provisional values for I_e and Bpl for different codecs are provided in [11]. A drawback of the random loss model used by ITU-T G.113 is the assumption of independent random losses which does not hold for many real networks such as VoIP and cellular [11]. Burst losses are also modeled using an equivalent random loss model, and are applicable for packet loss rates of $\leq$ 2% [11]. Furthermore, the equipment impairment factor and packet loss robustness factors have been evaluated for a very specific sample of burst packet loss, and may not reflect the impairment due to burst packet loss in general. Due to the above limitations, we observe that the E-model MOS values based on packet loss rates do not capture the variation in quality due to a distribution of the packet losses and as a result, the E-model values may not reflect the quality experienced by the user in real-time voice communication networks. We discuss this in more detail in Section VII.

MOS_x has the interpretation of voice outage rate that guarantees that the MOS is greater than MOS_x for x% of the realizations. It has a similar interpretation to outage capacity in the sense that increasing x would reduce the number of users that can be guranteed to achieve an MOS greater than MOS_x for a specified SNR. Analogously, increasing x for a specified MOS threshold would require a larger SNR to obtain the corresponding MOS. A higher value of MOS and x indicates good quality for a large percentage of the realizations and would require higher SNRs. Thus MOS_x is a critical indicator for voice quality performance in multipath fading. However, there are important differences between outage capacity and voice outage rate. Outage capacity is dependent on γ (SNR) and its probability distribution function whereas MOS_x takes into account the packet loss distribution over different realizations as well as the effect of packet loss concealment of G.711 and G.729 codecs. MOS_x can also be used to obtain link adaptation thresholds for switching between different rates specified in IEEE 802.11a.

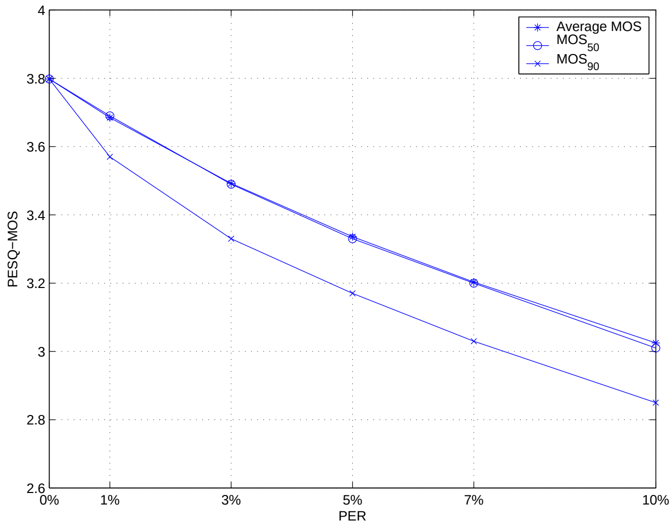

Fig. 3. Comparison of average MOS, MOS_{50} and MOS_{90}

V. Performance Evaluation in Multipath fading and AWGN channels

In this section we compare the performance of voice quality based on MOS_x for different choices of codecs, payload size and data rates. The complementary cumulative distribution functions (CCDF) of MOS for 10 ms G.711 payloads at 6 Mbps and 54 Mbps are plotted in Figs. 4 and 5, respectively. The voice quality performance is similar in both these cases with 54 Mbps operating at 15 - 35 dB and 6 Mbps operating at much lower SNRs of 0 - 15 dB.

Figure 6 plots the CCDFs for 20 ms G.729 payloads at a data rate of 6 Mbps. G.711 has a higher intrinsic MOS (4.4) compared to G.729 (3.8) while G.729 can support a higher number of voice users [2]. It is interesting to note that though the intrinsic MOS of 10 ms G.711 is higher than 20 ms G.729, the voice quality obtained at lower SNRs is quite comparable for both codecs. By comparing Figs. 4 and 6, we observe that for both 10 ms G.711 and 20 ms G.729 at the 6 Mbps data rate, an MOS of 3.0 is achieved for 36% of the realizations at 0 dB, and at an average SNR of 5 dB, MOS = 3.0 is achieved for 82% of the realizations. Though the intrinsic MOS of G.711 is better than G.729, the payload size of 10 ms G.711 is 80 bytes while the 20 ms payload of G.729 is 20 bytes and therefore, the larger payload size has a higher packet error rate in multipath fading scenarios. Hence, at lower SNRs in a multipath fading channel, both G.711 and G.729 have comparable performance. However, to obtain a higher quality and at

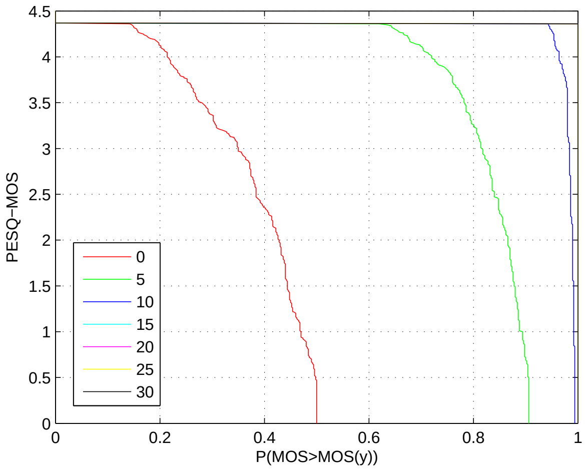

Fig. 4. CCDF for 10ms G.711 payload sizes at 6 Mbps for different SNRs in frequency selective multipath fading

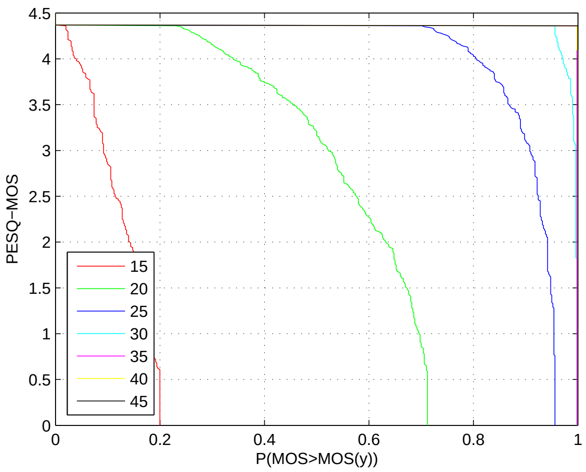

Fig. 5. CCDF for 10ms G.711 payload sizes at 54 Mbps for different SNRs in frequency selective multipath fading

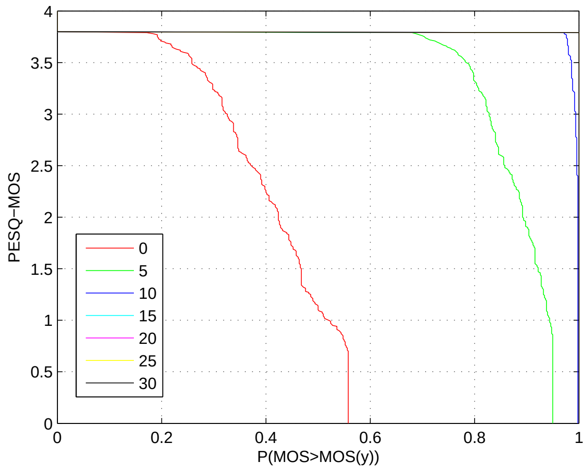

Fig. 6. CCDF for 20ms G.729 payload sizes at 6 Mbps for different SNRs in frequency selective multipath fading

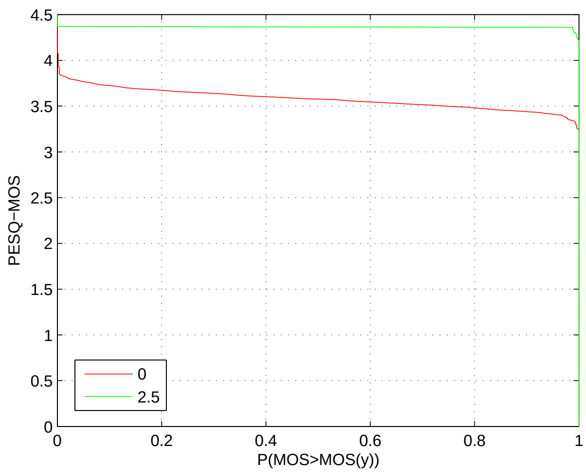

Fig. 7. CCDF for 10ms G.711 payload sizes at 6 Mbps for different SNRs in AWGN only channel

higher SNRs, G.711 is a better choice than G.729 due to its higher intrinsic MOS.

The CCDF plots for 10 ms G.711 in AWGN at 6 Mbps is plotted in Fig. 7. As expected, a higher SNR is required to get similar performance in multipath fading compared to AWGN only channels. For instance from Fig. 7, we observe that at 0 dB SNR an MOS greater than 3.5 can be achieved for 80% of the realizations while to achieve a similar performance in multipath fading (Fig. 4) would require an average SNR close to 7 dB.

VI. Link Adaptation

In IEEE 802.11a WLANs, as the channel is time varying in nature, the transmission parameters should be adapted according to channel conditions to improve link performance. The mechanism to select one of the multiple available transmission rates is referred to as *link adaptation*. The current link adaptation schemes used in IEEE 802.11a wireless cards are proprietary (mostly based on received signal strength and packet error rates) and in many cases can lead to inefficient bandwidth utilization and unnecessary rate adaptation. Most of the link adaptation criteria are based on maximizing throughput with a certain number of retransmissions [12]. In [13], we propose link adaptation schemes based on maximizing throughput with and without a packet error rate constraint in multipath fading channels.

We compare the link adaptation thresholds in IEEE 802.11a WLANs for different performance criteria. The link adaptation thresholds to maximize throughput are tabulated in Table I [13]. The link adaptation thresholds to maximize throughput with a packet error constraint are tabulated in Table II [13]. The link adaptation thresholds to maximize voice capacity while guaranteeing an MOS greater than 3.0 using 20 ms G.729 payloads with RoHC for 98% of the time is tabulated in Table III [2].

Using the E-model, for an average PLR of 5%, the MOS is specified as 3.3. However, by comparing Tables II and III, we observe that in order to achieve a MOS of only 3.0 for 98% of the realizations, significantly higher SNR is required than the scheme with a PLR constraint of 5%. Thus, we again observe that schemes using average MOS or average PLR for link adaptation would not be able to guarantee a satisfactory quality for a large percentage of the channel realizations. As a result, link adaptation based on MOS_x is a more reliable indicator of user satisfaction for voice transmission over wireless networks.

TABLE I

LINK ADAPTATION THRESHOLDS USING 20 BYTES PAYLOAD WITHOUT PACKET ERROR RATE CONSTRAINT IN A NORMALIZED MULTIPATH FADING ENVIRONMENT

Data Rate (Mbps)	6	12	24	36	48
SNR range (dB)	0-5	5-12	12-20	20-22	> 22

TABLE II

LINK ADAPTATION THRESHOLDS USING 20 BYTES PAYLOAD WITH AVERAGE PACKET ERROR RATE CONSTRAINT OF 5% IN A NORMALIZED MULTIPATH FADING ENVIRONMENT

Data Rate(Mbps)	6	12	18	24	36	48
SNR range(dB)	7-10	10-14	14-15	15-20	20-23	> 23

TABLE III

LINK ADAPTATION SCHEME FOR 802.11A USING G.729 UNDER MULTIPATH FADING WITH $Pr(MOS > 3.0) \geq 0.98$

Rate (Mbps)	6	12	18	24	36	48	54
SNR (dB)	9-14	14-18	18-19	19-24	24-27	27-29	>29

VII. CONCLUSIONS

We present a comprehensive evaluation of voice quality in frequency selective multipath fading channels for different combinations of codecs, payload size and data rates. We show that average SNR, packet loss rate, and average MOS are not good indicators for voice quality assessment since the average MOS is only achieved 50% of the time. We then define a new voice quality indicator, called MOS_x, which allows us to specify the MOS value obtained x% of the time. Unlike prior work, we use PESQ-MOS on coded speech with packet loss concealment. A comparison of the delivered voice quality over AWGN and multipath fading channels is provided and it is observed that realistic channel models are necessary to accurately assess the speech quality experienced by a user. We also propose a link adaptation scheme based on maximizing the voice capacity of a network with a MOS constraint.

REFERENCES

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

[2] N. Shetty, S. Choudhury, and J. Gibson, “Voice Capacity under Quality Constraints for IEEE 802.11a based WLANs,” *International Wireless Communications and Mobile Computing Conference*, July 2006.

[3] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Artech House, Jan 2000.

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

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

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

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

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

[9] 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.

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

[11] ITU-T G.113, “Transmission Impairments due to Speech Processing, Appendix 1: Provisional Planning Values for the Equipment Impairment Factor Ie, and Packet Loss Robustness Factor Bpl,” May 2002.

[12] D. Qiao, S. Choi, and K. G. Shin, “Goodput Analysis and Link Adaptation for IEEE 802.11a Wireless LANs,” *IEEE Trans. on Mobile Computing (TMC)*, vol. 1, no. 4, 2002.

[13] S. Choudhury and J. Gibson, “Joint PHY/MAC based link adaptation for wireless LANs with multipath fading,” *Wireless Communication and Networking Conference (WCNC) 06*, April 2006.