Capacity Evaluation of VoIP in IEEE 802.11e WLAN Environment

Abdelbasset Trad, Farukh Munir
INRIA, Planete Project
2004 Route des Lucioles, BP-93
06902 Sophia-Antipolis, France
Email: {atrad, mfmunir}@sophia.inria.fr

Hossam Afifi
INT-INRIA
9 rue Charles Fourier
91011 Evry, France
Email: Hossam.Afifi@int-evry.fr

***Abstract*— In this paper, we present an analytical model for VoIP capacity in IEEE 802.11e WLAN. We illustrate performance results relative to typical codec rates of G.711 PCM (64 kbit/s), G.729 (8 kbit/s) and G.723.1 (6.3 kbit/s). G.729 and G.723.1 allow a greater capacity than G.711 which is constrained by throughput. This greater capacity is at the expense of small quality degradation due to the delay increase since G.729 and G.723.1 codecs are more delay sensitive than G.711. In our study we analyse the occurrence of CAPs (Controlled Access Periods) during the Contention Period (CP) and its effect of a promising increase in the VoIP over WLAN capacity. We also show that high data rates (up to 54 Mb/s) allow important VoIP capacity (up to 400 G.711 VoIP calls, 997 G.729 VoIP calls and 1045 G.723.1 VoIP calls).**

I. Introduction

Wireless VoIP, typically over 802.11 WLAN, is becoming increasingly popular, but even further elevates the challenges of delay and loss reduction. Degradation of speech quality caused by packet delay and loss of voice traffic is still one of critical technical barriers of the VoIP system. Furthermore, apart from these limitations WLANs will need to support a large number of concurrent VoIP communications since VoIP is spreading rapidly especially in public spaces. These motivations led us to study the VoIP capacity in IEEE 802.11e WLAN and to investigate increasing this capacity by reducing VoIP codec rate while maintaining an overall good quality. A typical IEEE 802.11 WLAN with 11Mbps bandwidth could only support a very limited VoIP connections in DCF/PCF mode. In [14] the capacity of G.711 VoIP using CBR model and a 20 ms packetization interval was 12 calls. In this paper, we introduce an analytical VoIP capacity model in order to assess the VoIP capacity for the upgrade version of 802.11e MAC. In our study we analyze the occurrence of CAPs (Controlled Access Periods) during the CP (Contention Period) and its effect of a promising increase in the VoIP over WLAN capacity while keeping a low voice delay. We also show the benefit of recent technologies such as IEEE 802.11a with high data rates (up to 54 Mb/s), that will allow a high VoIP capacity. The paper is organized as follows: Section II overviews the VoIP system. In Section III, we investigate the limitations for the VoIP transmission over 802.11 networks and we analyze the 802.11e QoS enhancements. Section IV describes related work on VoIP over WLAN capacity evaluation. The studied network system is described in Section V. In Section VI, we introduce an analytical model for VoIP capacity over 802.11e networks. We show the effect of varying voice codec rate jointly with the durations of SF (SuperFrame) and CP on the number of simultaneously supported VoIP calls. Analytical modeling results show that the EDCA/HCCA mode of operation enhances the capacity of the network in terms of the number of simultaneous supported VoIP. Finally, Section VII concludes the paper.

II. VoIP System Overview

This section introduces a short description of VoIP basic mechanisms and the E-Model used for objective assessment of audio quality.

A. Voice Coding and Transmission over IP Networks

The commonly used VoIP codecs are G.711 [1], G.729 [2] and G.723.1 [3]. The traditional sample-based VoIP encoder G.711 uses Pulse Code Modulation (PCM) to generate 8 bits samples per 0.125 ms, leading to a data rate of 64 kb/s. Recent frame-based encoders can be used in order to provide drastic rate reduction (e.g., 8 kb/s for G.729, 5.3 and 6.3 kb/s for G.723.1). The reduced bandwidth utilization is at the expense of additional complexity and encoding delay as well as slightly lower quality. Further reduction in the data rate can be achieved by means of Voice Activity Detection (VAD), in which case no signal is encoded during silence periods. When silence suppression is employed, the codecs then operate in two states: a silent state at zero bit-rate and an active state at the compressed bit-rate. Regardless of the state, the frame period and frame size are still fixed. After the coding operation, the packetizer encapsulates a certain number of speech samples (for G.711) or a certain number of frames (for G.729 and G.723.1) into packets of equal sizes. The protocol stack used to carry the real time voice packets is RTP over UDP/IP. As the voice packets are sent over an IP network, they are subject to variable delays and network drops. Even if a lot of voice codecs can tolerate some small packet loss without severe degradation, voice traffic has unacceptable performance if long delays are incurred. It is recognized that the end-to-end delay has a great impact on the perceived quality of interactive conversations with a threshold effect around 150 ms [9]. For intra-continental calls, the packet delay is on the order of 30 ms and for inter-continental calls the delay can be as large as 100 ms [17]. The impact of delay on voice communication quality varies significantly with the use. For instance, long delays are not annoying in a cell phone as in a regular wired phone because of the added value of mobility.

B. The E-Model for Real-Time Measure of Audio Quality

Perceived voice quality is typically estimated by the subjective Mean Opinion Score (MOS), an arithmetic average of opinion score that ranges from 1 (unacceptable) to 5 (excellent). Objective quality scores can be generated by comparing the impaired voice signal with its original version such as in PESQ [4]. However, PESQ does not consider the effect of delay on voice communications and neither MOS nor PESQ can be used for real-time on-line quality estimation. The E-model is an analytic model defined in the ITU-T recommendation G.107 [5] as well as other associated ITU-T recommendations [7], [8], it provides a framework for an objective on-line quality estimation based on network performance measurements (e.g., delay and loss) and application level factors (e.g., low bit rate codecs). The result of the E-model is the calculation of the R-factor (ranging from a best case of 100 to a worst case of 0).

$$R = R_0 - I_s - I_d - I_e + A \tag{1}$$

Where R_0 groups the effects of noise, I_s includes the effect of other impairments related to the quantization of the voice signal, I_d represents the impairment caused by delay, I_e covers the impairments caused by low bit rate codecs and packet losses. The advantage factor A compensates for the above impairments under various user conditions. For mobile telephony A is assumed to be 10. We consider that A is 0 in the case of VoIP. The R-factor can be further translated into MOS scale through this expression:

$$MOS = 1 + 0.035R + 7.10^{-6}R(R-60)(100-R) \tag{2}$$

Table I shows quality classes and corresponding values of MOS and R-factor. The PSTN quality falls in the range 70

TABLE I
SPEECH TRANSMISSION QUALITY AND CORRESPONDENT MOS AND RATING FACTOR R VALUES

Speech Transmission Quality	MOS	R-factor
Best	4.50-4.34	100-90
High	4.34-4.03	90-80
Medium	4.03-3.60	80-70
Low	3.60-3.10	70-60
Poor	3.10-2.58	60-50

so that R = 70 (MOS = 3.6) will be our reference value for the capacity evaluation in this study.

III. Analysis of IEEE 802.11e QoS Enhancements

The original IEEE 802.11 standard [10] specifies two channel access mechanisms: a mandatory contention-based distributed coordination function (DCF) and an optional polling-based point coordination function (PCF). DCF provides a best effort service and is not capable of providing differentiation and prioritization based upon traffic type. While DCF may provide satisfactory performance in delivering best-effort traffic, it lacks the support for QoS requirements posed by real time traffic, and especially VoIP which has stringent delay requirements. These requirements make the DCF scheme an infeasible option to support QoS for VoIP traffic. PCF mode, with a centralized controller, represented another promising alternative to providing QoS in WLAN [13]. Nevertheless, studies on carrying VoIP over WLAN in PCF mode in [16] found that when the number of stations in a basic service set (BSS) is large, the polling overhead is high and results in excessive end-to-end delay and that VoIP still gets poor performance under heavy load conditions. Thus, neither DCF nor PCF presents sufficient functionality to provide the QoS demanded by multimedia applications. In WLANs, the physical layer's error rate is larger than that of wired LAN and the challenges of the wireless channel make physical layer data rate improvements difficult to achieve. These reasons have led to the development of service differentiation based MAC schemes that classify traffic types based on their relative priorities. The IEEE 802.11e standard [11] addresses the shortcomings of the 802.11 standard, defines a superset of features backward compatible with DCF/PCF and introduces the Hybrid Coordination Function (HCF). HCF has two modes of operation: Enhanced Distributed Coordinated Access (EDCA) and HCF Controlled Channel Access (HCCA). EDCA inherits all the contention schemes and parameters of the original 802.11 DCF and provides service differentiation through prioritized access to the wireless medium. Prioritization is realized through the introduction of four Access Categories (ACs) each with its own transmit queue and set of AC parameters. The differentiation in priority between ACs is realized by setting different values for the AC parameters which include the arbitration interframe spacing (AIFS) and the minimum contention window size (CWmin). With proper tuning of these parameters, the performance of delay sensitive multimedia traffic can be improved. In an infrastructure network, the AP will gain access to the medium with a higher priority than other QoS Stations (QSTAs). Under HCF the basic unit of allocation of the right to transmit onto the wireless medium is the transmission opportunity (TXOP). Each TXOP is defined by a starting time and a defined maximum length. The TXOP may be obtained by a QSTA winning an instance of EDCA contention during the CP, or by a non-AP QSTA receiving a QoS + CF-Poll during the CP or CFP. With the HCCA, a hybrid coordinator (HC) allocates transmission opportunities (TXOPs) to wireless STAs by polling, so as to allow them contention-free transfers of data, based on QoS policies. QSTAs may obtain TXOPs using one or both of the channel access mechanisms. An HC generates an alternation of CFP and CP, the sum of the two periods forms the "superframe" (SF). In addition, contrary to DCF, QSTAs can be polled during the CP in periods called Controlled Access Periods (CAPs) as shown in Figure 1. The duration of TXOPs allocated to each QSTA is determined by the HC scheduler according to the requested QoS parameters. It is also possible to change the SF length since beacons carry a parameter indicating the SF length. Due to the service differentiation, the real-time traffic gets a higher priority in winning channel contention under 802.11e and evidently provides a better performance as compared to the basic 802.11 MAC scheme.

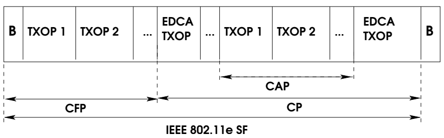

Fig. 1. Alternation of CFP, CP and triggered CAP during an IEEE 802.11e superframe

IV. Related Work

There is a rich literature on the capacity evaluation of WLANs in terms of VoIP calls. In [12], simulation results show that, using HCF mode of operation, the number of simultaneous transmissions can be increased to a varying extent because of the 802.11e service differentiation. [14] determines an analytical upper bound of the capacity for VoIP calls in IEEE 802.11b networks. In that paper, many scenarios were analytically studied including wireless channel conditions, voice coding schemes and different delay constraints. The capacity of G.711 VoIP using CBR model and a 20 ms packetization interval was 12 calls. In [15], the authors show that the capacity of VoIP in 802.11a can be improved by using automatic rate selection instead of a fixed 6Mbp/s physical link rate. In [13], the authors analyze the capacity of a system that uses PCF for VoIP traffic, CBR and VBR models were considered. Values of 75 ms and 90 ms were used as CFP interval. The capacity for VoIP with a 90 ms CFP was 26 voice calls, but the maximum delay was 303 ms. This end-to-end delay value is not acceptable for an excellent quality of voice. Our work is different in considering an analytical VoIP capacity model for the upgrade version of 802.11e MAC in order to assess the improvements in VoIP performance compared to the basic 802.11 MAC scheme. In our study we analyse the occurrence of CAPs during the CP and its effect of a promising increase in the VoIP over WLAN capacity while keeping a low voice delay.

V. Studied Network System

We consider the architecture presented in Figure 2 where a VoIP gateway acts as the conjuncture of wired Internet and wireless LAN. The gateway is associated with an 802.11e QoS AP (QAP). A QAP is required to support VoIP calls between wired and wireless networks. In such a situation, the functionality of HC is performed at the QAP. We assume that all VoIP packets are exchanged via the QAP. Thus, for both VoIP calls between two QSTAs in the WLAN (intra-AP call) and calls to the wired network, sending and receiving ends are polled by the QAP. In the case of a call between a QSTA and a wired station, we consider that the VoIP gateway behaves as a wireless end. The wired network effect on VoIP call quality is not studied in this paper. For a voice call between two QSTAs, the two ends can directly send traffic to each other without further intervention of the AP. We will not consider this case because it is an implementation concern and it is optionally provided.

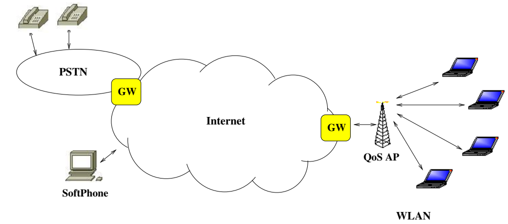

Fig. 2. Network Architecture

VI. VoIP Capacity in 802.11e Networks

The WLAN network capacity in terms of VoIP calls is defined as the maximum number of bi-directional calls that a given AP can support while maintaining acceptable level of voice transmission quality. In this study, we use Constant Bit Rate (CBR) model for VoIP traffic. Parameters used for this modelling study are described in Table II.

TABLE II
System Modelling Parameters

B	Beacon frame size (320 bits)
H	Protocol header overhead (456 bits)
P	Physical layer header (192 bits)
T_B	Beacon interval (0.01 sec)
$SIFS$	SIFS time interval (28μs)
CF_{end}	CFP end frame (192 bits)

After the acceptance of a voice call by the called destination (wireless station or VoIP gateway), the QAP adds the two ends of the call to its polling list. We assume that all the voice data generated during the interpoll time will be sent in one packet in order to reduce the protocol header overhead. Thus, the number of packets generated from different VoIP communications that can be sent during one packetization interval is the maximum number of calls. In order to find the maximum number of active stations that can be supported simultaneously for real-time CBR VoIP traffic, we consider that the maximum interpoll time of VoIP calls is the duration of the SF. This statement should be satisfied in order to avoid queue build up at each node. In CP, the Controlled Access Period (CAP) will be started whenever the AP will find the channel busy for a PIFS interval, and the probability to have a station finished its packetization before the arrival of a poll from the AP is quite high in this case. Therefore, we consider two different poll periods in our system during the CFP (T_{CFP}) and the CP (T_{CP}) and the total number of simultaneously supported VoIP calls would be an addition of the calls supported in the two separate poll periods. In order to define the packet size, assume C to be the codec rate at which the codec is generating packets, T_{voice} is the packetization delay (listed in Table III), T_{CFP} is the maximum interpoll time to voice call in the CFP. So the largest possible packet time interval for a bi-directional voice call polled during the CFP will be:

$$T_{poll-cfp} = \frac{(C * (T_{voice} + T_{CFP} + H + P) * 2)}{R} + 2 * SIFS \quad (3)$$

Where R is the wireless channel transmission rate, SIFS is the time interval used to separate between all fragments and ACKs. Similarly, we obtain the equation for the maximum packet size of voice calls polled during the Controlled Access Period (CAP) of the CP:

$$T_{poll-cp} = \frac{(C * (T_{voice} + T_{CP} + H + P) * 2)}{R} + 2 * SIFS \quad (4)$$

The total number of supported voice calls can then be obtained from the sum of the two equations 5 and 6, presenting the voice calls during CFP and CP respectively:

$$N_{poll-cfp} = \frac{T_{CFP} - T_{overhead}}{T_{poll-cfp}} \quad (5)$$

$$N_{poll-cp} = \frac{T_{CP}}{T_{poll-cp}} \quad (6)$$

The overhead time related to the CFP is obtained from [13] and is given by the equation 7:

$$T_{overhead} = (\frac{B + P}{R} + SIFS) * (\frac{SF - T_{CP}}{T_B}) + \frac{CF_{end} + P}{R} \quad (7)$$

We compare the capacity of CBR VoIP using G.711, G.729 and G.723.1 standard voice codecs, features of these codecs are listed in Table III. Figure 3 compares the quality of these codecs using their MOS as function of increasing network delay and assuming no packet loss. From this figure, we observe that MOS value of 3.6 (PSTN reference quality) can be obtained for the three codecs G.723.1, G.729 and G.711 for delay values under 170 ms, 240 ms and 325 ms respectively.

TABLE III

CODEC BIT RATE AND PACKETIZATION INTERVALL

Codec	Bit Rate (Kbit/s)	Packetization Delay (ms)
G.711	64	20
G.729	8	20
G.723.1	6.3	30

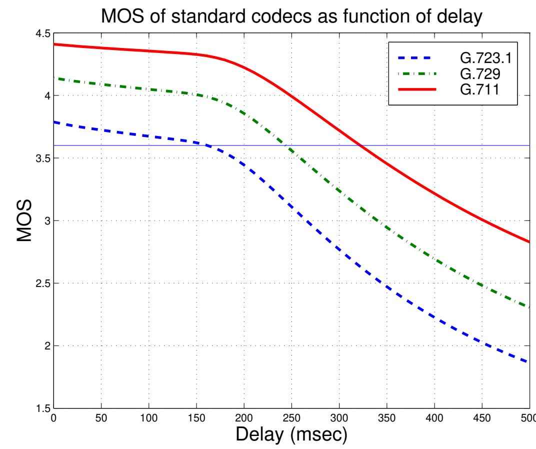

Fig. 3. MOS of typical VoIP codecs as function of network delay

As opposed to the previous studies [13][14], we consider that SF length can be modified because of the extension of voice calls being polled during the CP (in CAPs) along with the data traffic contending to win the channel. In Figure 4, we plot, for three different coding rates, VoIP capacity in an IEEE 802.11 network (11Mbps) using DCF/PCF mode of operation (no polling during the CP). The SF size varies from 30 ms to 120 ms. The maximum number of VoIP calls, with a 120 ms SF, is of 20 calls when the G.711 codec is used and is much higher for G.729 and G.723.1 (95 and 107 respectively). This is because G.711 is constrained by throughput. Using G.729 and G.723.1 low bit rate codecs improve the WLAN capacity, but at the expense of small quality degradation due to the delay increase. Quality of G.729 and G.723.1 is more sensitive to delay than G.711 as shown in Figure 3. Figure 5 illustrates the VoIP capacity in a 802.11e WLAN, using the analytical model that we proposed. Under EDCA/HCCA mode, a large increase in the VoIP capacity is achieved. For a 120 ms SF, the maximum number of G.711 VoIP calls is 99 (20 calls with DCF/PCF). G.723.1 VoIP capacity reaches

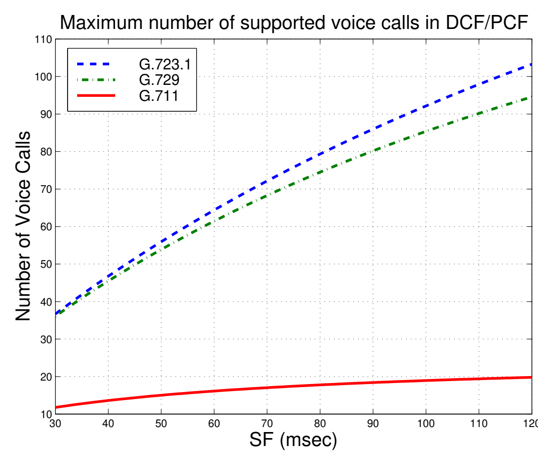

Fig. 4. Maximum number of supported voice calls in DCF/PCF

392 calls (107 calls with DCF/PCF). This capacity increase

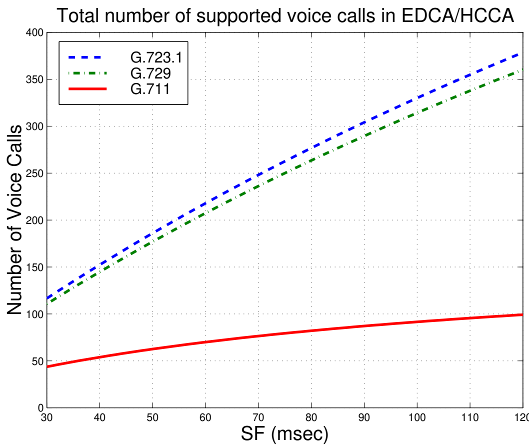

Fig. 5. Improved maximum number of supported voice calls in EDCA/HCCA

is due to the extension of Controlled Access Periods (CAP) during the CP mode of operation. During CP, QSTAs have to contend for the channel and certain duration of time is wasted in contention winning and backoff window exhaustion process. Whereas in the CAP the Hybrid Coordinator (HC) located at the QAP gains control of the medium after sensing the channel idle for PIFS duration. Then the AP polls the QSTAs which has voice traffic based on its defined polling scheme. In this fashion, when the CFP is terminated (we consider non-empty queues), the voice stations are not required to wait for the next occurrence of CFP. The wireless stations can get another poll from the HC in the CP when an occurrence of CAP in undergoing, providing low end-to-end delay for voice stations. The VoIP capacity is also improved when the SF duration is increased, but this will cause an additional delay.

In none of the earlier works, the effect of the duration of CP and CFP in a SF is directly addressed to show the difference in the number of voice calls. In Figure 6, we present the VoIP capacity results for the three before-mentioned codecs while increasing the CP duration from 30 ms to 120 ms (SF size is kept constant: 120 ms).

The maximum VoIP capacity is obtained for CP values between 50 ms and 60 ms. From this result, we conclude that simply by tuning the duration for CP and CFP out of superframe duration, the number of voice calls being supported simultaneously can be increased due to existence of CAP during which voice stations are polled by the QAP. Figures 7 and 8 illustrate the effect of the physical transmission rate on the VoIP capacity for DCF/PCF and EDCA/HCCA respectively.

Table IV and V summarize numerical results of the capacity according to standard physical data rates. Recent technologies such as IEEE 802.11a with high data rates (up to 54 Mb/s) promise important VoIP capacity (up to 400 G.711 VoIP calls, 997 G.719 VoIP calls and 1045 G.723.1 VoIP calls) when the EDCA/HCCA mode is used. For lower data rate (e.g., 11Mbps of the IEEE 802.11b standard) VoIP capacity is not so high,

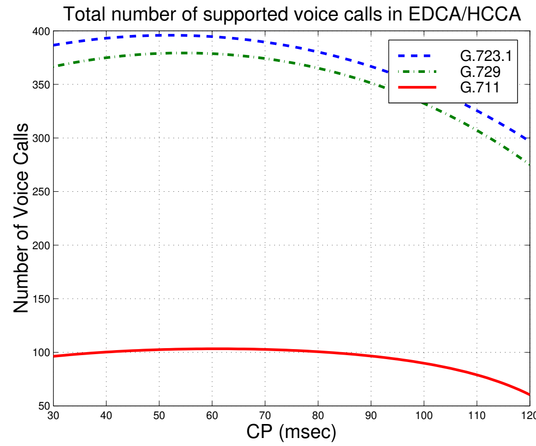

Fig. 6. Maximum number of supported Voice calls while increasing CP duration in EDCA/HCCA mode of function

especially for G.711 codec (99 G.711 VoIP calls for 11Mbps data rate) which is required in order to ensure an excellent voice quality.

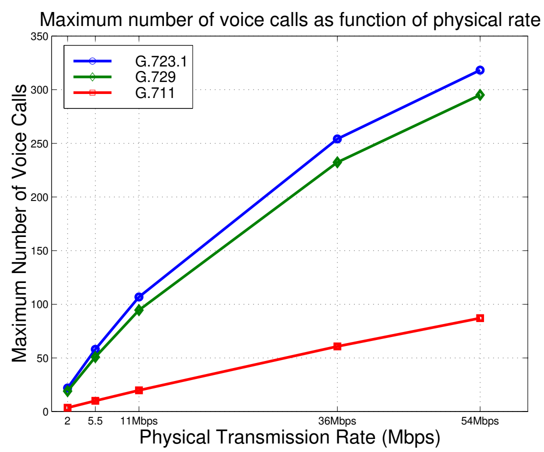

Fig. 7. Maximum number of supported voice calls in DCF/PCF with a SF of $120ms$ as function of increasing physical transmission rate

TABLE IV

MAXIMUM NUMBER OF VOIP CALLS FOR STANDARD VOICE CODECS AS FUNCTION OF CHANNEL TRANSMISSION RATE IN DCF/PCF MODE

Physical Rate	G.711 (64kb/s)	G.729 (8kb/s)	G.723.1 (6.3kb/s)
2 Mbps	4	19	46
5.5 Mbps	10	51	58
11 Mbps	20	95	107
36 Mbps	61	232	254
54 Mbps	87	295	318

VII. CONCLUSION

In this paper, we have proposed an analytical model for VoIP capacity in IEEE 802.11e WLAN. Modelling results show that

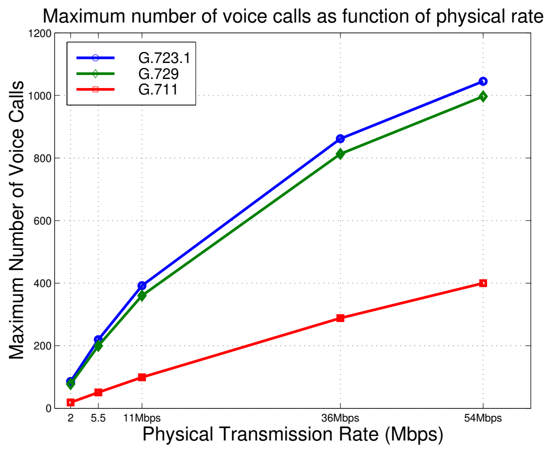

Fig. 8. Maximum number of supported voice calls in EDCA/HCCA with a SF of $120msec$ as function of increasing physical transmission rate

TABLE V

MAXIMUM NUMBER OF VOIP CALLS FOR STANDARD VOICE CODECS AS FUNCTION OF CHANNEL TRANSMISSION RATE IN EDCA/HCCA MODE

Physical Rate	G.711 (64kb/s)	G.729 (8kb/s)	G.723.1 (6.3kb/s)
2 Mbps	19	77	86
5.5 Mbps	51	199	219
11 Mbps	99	360	392
36 Mbps	288	813	861
54 Mbps	400	997	1045

the EDCA/HCCA mode of operation enhance the capacity of the network in terms of the number of simultaneous supported VoIP calls. This capacity increase is due to the extension of Controlled Access Periods (CAP) during the CP mode of operation. We compared the capacity of CBR VoIP using G.711, G.729 and G.723.1 standard voice codecs. Low bit rate codecs (i.e., G.729 and G.723.1) provide higher VoIP capacity, at the expense of small quality reduction. We also observed that simply by tuning the duration for SF and CP duration, the number of voice calls being supported simultaneously can be increased due to the existence of CAP during which voice stations are polled by the QAP. In future work, we will study the effect of packet loss on the VoIP capacity and model the delay of voice packets as function of the number of carried calls.

REFERENCES

[1] ITU-T Recommendation G.711. *Pulse Code Modulation (PCM) of voice frequencies.* ITU-T, November 1988.

[2] ITU-T Recommendation G.729 Annex A. *C source code and test vectors for implementation verification of the G.729 reduced complexity 8 kbit/s CS-ACELP speech codec.* ITU-T, November 1996.

[3] ITU-T Recommendation G.723.1 *Speech coders : Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s.* ITU-T, March 1996.

[4] ITU-T Recommendation P.862 *Perceptual evaluation of speech quality (PESQ): An objective method of end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.* ITU-T January 2001.

[5] ITU-T Recommendation G.107. *The Emodel, a computational model for use in transmission planning.* ITU-T, March 2005.

[6] ITU-T Recommendation G.108. *Application of the Emodel: a planning guide.* ITU-T, September 1998.

[7] ITU-T Recommendation G.109. *Definition of categories of speech transmission quality.* ITU-T, September 1999.

[8] ITU-T Recommendation G.113. *Transmission impairments due to speech processing.* ITU-T, February 2001.

[9] ITU-T Recommendation G.114. *One way transmission time.* ITU-T, May 2003.

[10] IEEE Std 802.11-1999, Part 11: *Wireless LAN MAC and physical layer specifications.* Reference number ISO/IEC 8802-11:1999(E).

[11] IEEE 802.11e/D9.0. *Wireless MAC and physical layer specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements.* August 2004.

[12] A. Grilo and M. Nunes *Performance Evaluation of 802.11e.* In Proceedings of IEEE PIMRC 2002.

[13] M. Veeraraghavan, N. Cocker, and T. Moors. *Support of voice services in IEEE 802.11 Wireless LANs.* In Proceedings of IEEE INFOCOM 2001, Alaska, April 2001.

[14] D. P. Hole and F. A. Tobagi. *Capacity of IEEE 802.11b Wireless LAN supporting VoIP.* In Proceedings of IEEE ICC 2004, Paris, June 2004.

[15] N. Smavatkul, Y. Chen, and S. Emeott. *Voice capacity evaluation of IEEE 802.11a with automatic rate selection.* In GLOBECOM 2003, San Francisco, December 2003.

[16] D.-Y. Chen, S. Garg, M. Kappes, and K. S. Trivedi. *Supporting VBR VoIP traffic with IEEE 802.11 WLAN in PCF mode.* In Proceedings of OPNETWork 2002, Washington D. C., August 2002.

[17] M. Karam and F. Tobagi. *Analysis of the Delay and Jitter of Voice Traffic over the Internet.* In Proceedings of IEEE INFOCOM 2001, Anchorage, AL, April 2001.