

Voice over IEEE 802.11b Capacity

M. Coupechoux, V. Kumar, and L. Brignol

Abstract—In this paper, we study the VoIP capacity of IEEE 802.11b (DCF), i.e., the maximum number of simultaneous voice calls that can take place in a WLAN cell. This capacity is highly dependent on the chosen codec and on the distance from the access point (AP) to the users. Thus several codecs are studied, namely G711, GSM-EFR, and G723.1. All users are assumed to be at a fixed distance from the AP. The quality of voice calls is evaluated thanks to the E-model. Simulation results provide the following results: G711 allows up to 5 simultaneous calls, GSM-EFR up to 12, and G723.1 up to 18 for a voice quality greater than $R = 70$ and a network delay of 20ms.

I. INTRODUCTION

The high data rate WLAN standard IEEE 802.11b is experiencing a successful development. This success is drawn by two main types of networks: enterprise networks and hot spots, i.e., conference centers, railway stations, airports, hotels, etc. In both environments, VoIP and VoWLAN are becoming attractive technologies with the main goal to reduce the communication costs by merging data and voice networks.

In this context, this paper provides the VoIP capacity of IEEE 802.11b, i.e., the maximum number of voice calls that can simultaneously take place in a WLAN cell. Providing such a real-time application over a CSMA/CA based network, firstly designed for best-effort traffic is a challenging issue. Numerous papers try to address this problem. [2] evaluates the capacity with the Point Coordination Function (PCF) of the standard. This feature is however optional and most of the card manufacturers didn't implement it in commercial products. Instead, the Distributed Coordination Function (DCF) is always used in practice. Some papers, e.g. [3] or [7], propose adaptations of DCF to allow voice traffic. [5] compares DCF, PCF, priority queuing and blackburst mechanisms.

Several experimental results based on IEEE 802.11b DCF are also available in the literature, e.g.

[6] [4] or [8]. [1] provides an analysis of the number of VoIP calls for different codecs and data rates. However, none of these studies bases its conclusions on an efficient model for voice quality. All cited papers use as metrics the packet loss and the average packet delay.

In this paper, conclusions are built on the E-model that is an efficient tool to predict the voice quality (section II). The system description is given in section III. This includes the chosen codecs, the traffic model, the MAC protocol, the channel model, and the link adaptation strategy. At last, section IV provides simulation results for the capacity as a function of the chosen codec and of the distance of the users to the AP. The influence of the mechanism implemented in the de-jittering buffer is also shown.

II. E-MODEL

A. Description

The E-model is a tool to predict how an “average user” would rate the voice quality of a phone call [14]. This model has been standardized by the ITU [10][11] and provides a R -scale. The rating factor R is composed of several additive terms, each one representing a specific source of voice quality degradation: $R = R_0 - I_s - I_d - I_e + A$.

R_0 is usually set to 94.3 and represents the basic signal-to-noise ratio. I_s represents impairments simultaneously occurring with the voice signal (e.g. quantization). I_d represents impairments due to transmission delays. I_e represents impairments caused by the use of a specific equipment, e.g., I_e is affected by the choice of the codec and by packet loss. A is the expectation factor, it represents the degradation that a user is likely to accept because he is aware that the technology is wireless and mobile.

The range of R is from the worst quality, 0, to the best one, 100. The quality classes are shown in Tab.I. Note that the PSTN quality falls in the range 70–100, so that $R = 70$ will be our cut-off value for the capacity evaluation.

M. Coupechoux is with Institut Eurecom, Sophia-Antipolis, France and with Alcatel Research and Innovation, Marcoussis, France (e-mail: marceau.coupechoux@alcatel.fr). V. Kumar and L. Brignol are with Alcatel Research and Innovation.

TABLE I
QUALITY CLASSES ACCORDING TO THE E-MODEL.

R range	90-100	80-90	70-80	60-70	0-60
Quality	best	high	medium	low	poor

The fine-tuning of the parameters of the E-model is important to get accurate results. For our simulations, we chose the values given in Tab.IV in Appendix. These are the default values given by the standard.

B. Mouth-to-ear Delay Budget

One of the main source of quality degradation is the mouth-to-ear delay. In this section, details of this delay are given. The main sources of delays are the following:

- The packetization time T_{pack} , i.e., the time needed to collect all voice samples that form a packet. In our simulations, each voice frame is packetized in a single IP packet, so that $T_{pack} = T_F$, where T_F is the voice frame duration.
- The voice encoding and decoding process T_{DSP} , i.e., the time needed to encode the analog voice source or to decode the voice samples to an analog signal. According to [16], $T_{DSP} = 12,5\text{ms}$ for the codec G729. We will assume that this value is the same for other considered codecs.
- The look ahead delay T_{LA} if any. Some codecs need indeed to collect a few samples before producing a voice frame [15].
- The network delay T_{nw} , i.e., the delay caused by the transmissions in the wired network and by the different buffers in traversed IP routers. In our simulations, two fixed values have been chosen: $T_{nw} = 20\text{ms}$ and $T_{nw} = 100\text{ms}$. These values are in accordance with those given in [17] and [18], and represents two extreme cases.
- The access delay T_{WLAN} , i.e., the queuing delay in the AP and the delay added by the MAC layer of IEEE 802.11. This delay will be simulated.
- The dejittering delay T_{jitt} , i.e., the delay introduced by the dejitterization buffer at the receiver side. The computation of T_{jitt} is detailed in the next section.

Hence, the overall mouth-to-ear delay is given by:

$$T = T_{pack} + T_{DSP} + T_{LA} + T_{nw} + T_{WLAN} + T_{jitt} . \quad (1)$$

Note that the G114 recommendation of the ITU-T [12] specifies the following upper bounds for the delay:

- $T \leq 150\text{ms}$: most applications are not significantly affected.
- $150 < T \leq 400\text{ms}$: this is still an acceptable delay, in particular for international calls.
- $T > 400\text{ms}$: this is an unacceptable delay.

C. Dejittering Mechanisms

In principle, the receiver of a voice call could play out the first packet as soon as it arrives in its reception buffer. Then, following packets have to be played out with a strict period in order to reproduce the streamed information. In practice, packet based networks introduce transport delay variations (or jitter), so that slowest packets can be lost, when their turn to be played out occurs.

That is the reason why the receiver has a dejittering buffer that retains fast packets until they have to be played out. The buffer “absorbs” the delay variations.

Let $s_n = nT_s$ be the sending instant of the n -th packet, where T_s is the sending period of the voice frames. Let d_n be its delay and $a_n = s_n + d_n$ its arrival instant at the receiver. We assume that the packets arrive in correct order. Routes in the IP networks are indeed very stable. Moreover, in the WLAN network, the “stop-and-wait” acknowledgement policy ensures the correct order.

The dejittering buffer retains the first packet for a time D . Then the buffer is read at periodical instants, i.e., $a_0 + D + nT_s$. If the n -th packet is present in the buffer at this time, it is played out. Otherwise, i.e., if it is too late, the packet is lost. This occurs if:

$$a_n > a_0 + D + nT_s \quad (2)$$

$$nT_s + d_n > s_0 + d_0 + D + nT_s \quad (3)$$

$$d_n > d_0 + D \quad (4)$$

$$D < d_n - d_0 \quad (5)$$

If the receiver doesn’t want any loss of packets, the dejittering delay has to be chosen as follows:

$$T_{jitt} = d_{max} - d_0 , \quad (6)$$

where d_{max} is the maximum delay. However, the voice traffic can tolerate some packet loss without a

big degradation of the quality. If P_{loss} is tolerated in the dejittering buffer, T_{jitt} is now:

$$T_{jitt} = d_q(P_{loss}) - d_0, \quad (7)$$

where $d_q(P_{loss})$ is the $(1 - P_{loss})$ -quantile of the delay. At this point, a trade-off has to be found because increasing P_{loss} reduces T_{jitt} and so the mouth-to-ear delay.

In practice, the receiver doesn't have the probability density function (pdf) of the delay. Several adaptive algorithms are presented in [19]. In this paper, a perfect mechanism is assumed that is able to obtain Eq.7.

D. Packet Losses

There are two main sources of packet loss. The first one is due to the MAC layer. If a packet is not correctly received because of the channel conditions or because of a collision, the MAC layer retransmits the lost packet. After 4 unsuccessful retransmissions of a RTS (Ready To Receive) or after 7 unsuccessful retransmissions of a data packet, the packet is definitely lost. The proportion of such packets is P_{loss}^{MAC} .

The second reason is due to the dejittering mechanism implemented at the receiver, as shown in the previous section. Among the received packets, the proportion of such packets is P_{loss} .

A third source of loss could be considered: the congestion in one of the nodes of the wired network, including the AP. But it is not taken into account in this paper.

III. SYSTEM DESCRIPTION

In this section, a description of the system including the network topology, the studied codecs, the traffic model, the channel model, and the link adaptation strategy, is provided.

A. Topology

Since the IEEE 802.11b includes a rate (or link) adaptation mechanism able to switch between the physical modes 1, 2, 5.5, and 11Mbps, the capacity of the cell depends on the spatial repartition of the users. In this paper, all users are assumed to be at an equal distance to the AP, as shown on Fig.1 for a distance d to the AP and seven terminals.

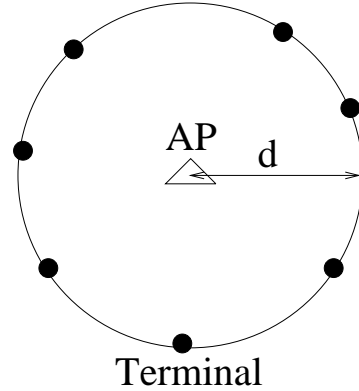


Fig. 1. Network topology with seven terminals.

B. Codecs and traffic model

Three codecs are considered: GSM-EFR, G711, and G723.1. Their main characteristics are summarized in Tab.II. Note that the value I_e is given in the case of no packet loss.

TABLE II
CODECS MAIN CHARACTERISTICS.

	GSM-EFR	G711	G723.1
Bit rate [Kbps]	12.2	64	6.3
Packet size [bits]	244	640	189
Frame duration [ms]	20	10	30
Look ahead [ms]	0	0	7.5
I_e	5	0	15

The dependence of I_e with packet loss is provided by [11]. For G711, the standard distinguishes G711 with and without Packet Loss Concealment (PLC). PLC increases the robustness against packet loss. Both cases have been considered assuming bursty packet loss.

Voice frames are sent in a IP/UDP/RTP packet. For these protocols, there is an overhead of $20 + 8 + 12 = 40$ bytes. Note that the physical header of IEEE 802.11b (with long preamble) adds 24bytes, and the MAC header adds 34bytes.

The voice traffic is modeled by a ON/OFF source in each direction. The mean ON period duration is 1.0 s, and the mean OFF period duration 1.35 s. Both follow an exponential distribution. Thus, the voice activity is 42.6%. These values are in accordance with [13].

C. Channel Model and Link Adaptation

Four physical modes have been defined by the standard IEEE 802.11b, namely 1, 2, 5.5, and 11Mbps. 1 and 2Mbps belongs to the basic rate set, i.e., all terminals must be able to receive and transmit at these data rates. In our implementation, the RTS (Ready To Receive)/CTS (Clear To Send) handshake is used for each packet. RTS, CTS, and ACK (Acknowledgement) control packet are transmitted at the basic data rate 1Mbps. The physical mode of data packet is chosen according to the link adaptation strategy. This mechanism is based on the channel quality. So, let us first describe the considered channel model.

The path loss is given by the following formula:

$$L = 32.4 + 20 \log(f) + 10n \log(d), \quad (8)$$

where $f = 2.4GHz$ is the frequency in GHz, $n = 4$, and d is the distance in m between the sender and the receiver. This model corresponds to an indoor propagation model. The received power is computed from link budget calculations with an additive log-normal distribution modeling shadowing (4dB).

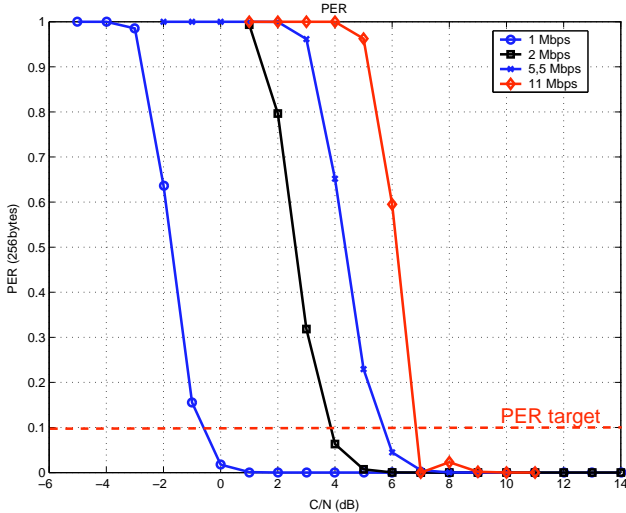


Fig. 2. PER vs. C/N for the four physical modes of IEEE 802.11b.

The packet error rate (PER) is approximated according to the received power. Errors on bits are assumed to be independent, so that the PER can be deduced from the bit error rate (BER) with the following formula: $PER = 1 - (1 - BER)^N$, where N is the number of bits in the considered packet. Then, the BER is computed by the analytical formulas of DBPSK, DQPSK, and the MBOK (CCK is considered as a variation of MBOK) modulations [20].

Fig.2 shows the PER performance of the four physical modes as a function of the carrier-to-noise ratio (C/N).

The policy of link adaptation employed in our simulations is based on PER metric and C/N switching thresholds. That means that the final decision for the physical mode takes into account both the received power C (C have to be above the sensitivity threshold of the mode) and C/N measurements (with PER constraint).

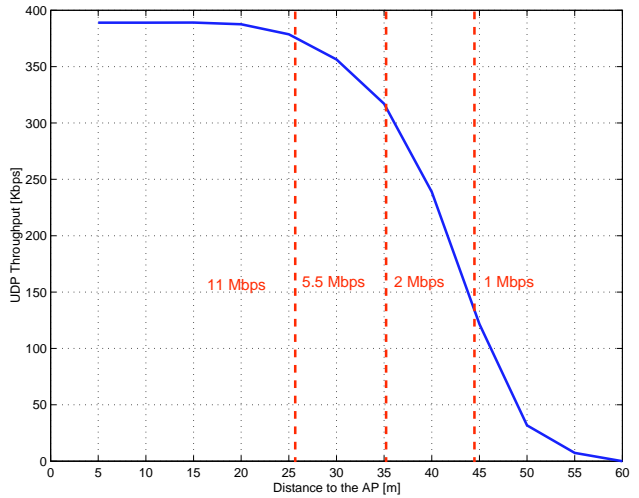


Fig. 3. UDP throughput vs. distance to the AP, packet load = 80bytes.

If the link adaptation has been designed for 256 bytes packets, and the PER target has been fixed to 0.1, C/N thresholds can be derived from Fig.2 (in dB): [3.8; 5.7; 6.7]. With these assumptions, the UDP throughput from the AP to a single terminal in the cell is presented on Fig.3 for 80 bytes packets. The ranges of the physical modes are also shown, based on the sensitivity thresholds. Note that some assumptions are quite optimistic. However, the computed range for an indoor environment is in accordance with experimental ranges.

IV. SIMULATION RESULTS

In this section, the aforementioned system is simulated thanks to the Network Simulator ns2 [21]. Capacity values are deduced from the E-model for different codecs, distances to the AP, and P_{loss} values. The simulated time is 200s and the voice quality is evaluated at the reception of one terminal.

A. Influence of the distance to the AP

Let us first have a look at the influence of the distance to the AP on the cell capacity. In this section,

$T_{nw} = 100\text{ms}$. As the distance from the AP to the terminals increases, the link adaptation mechanism degrades the physical mode. As a consequence, the available throughput above the MAC layer is reduced and less simultaneous voice calls are possible. In this section, $T_{nw} = 100\text{ms}$.

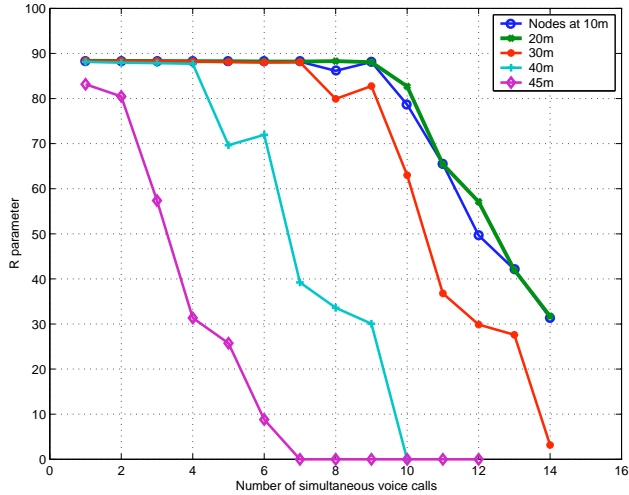


Fig. 4. R parameter vs. number of simultaneous voice calls, influence of the distance to the AP, GSM-EFR.

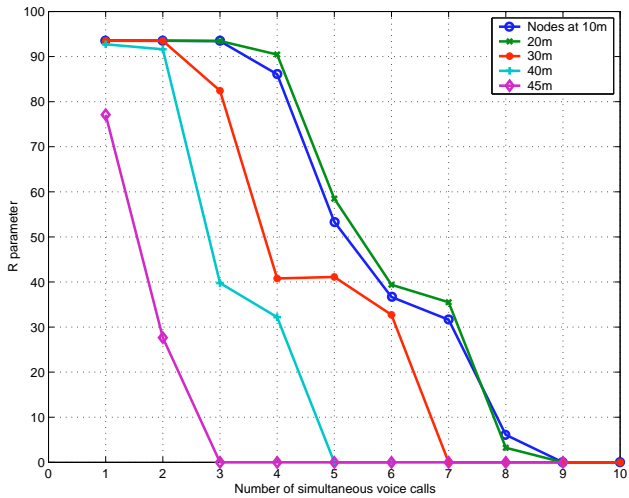


Fig. 5. R parameter vs. number of simultaneous voice calls, influence of the distance to the AP, G711.

The performance of the codecs GSM-EFR, G711, G711 with PLC, and G723.1 are shown respectively on Fig.4, 5, 6, and 7 as a function of the distance to the AP. Note that in this case $P_{loss} = 0\%$ (see Eq.7). This explains the small difference between G711 and G711 PLC.

With all codecs, there is a high degradation of the capacity with distance. With GSM-EFR, going from 10 m to 45 m reduces the capacity from 10 voice calls

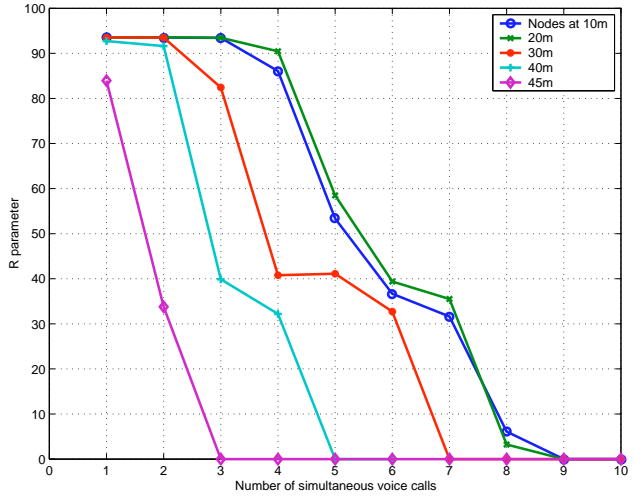


Fig. 6. R parameter vs. number of simultaneous voice calls, influence of the distance to the AP, G711 with PLC.

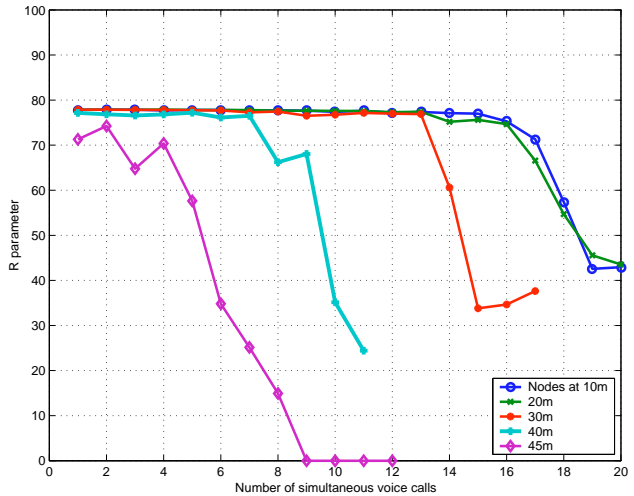


Fig. 7. R parameter vs. number of simultaneous voice calls, influence of the distance to the AP, G723.1.

to 2. With G711, while 4 simultaneous calls are possible at 10 m, a single call can be made at 45 m. The biggest degradation can be seen with G723.1, from 17 calls at 10 m down to 4 at 45 m.

These values can be surprising with respect to the available physical data rate available in the cell, especially at 10 m, where the physical mode is 11Mbps. In fact, IEEE 802.11b suffers from a huge overhead, due to the RTS/CTS handshake, the acknowledgment, the MAC header, the backoff window, and the basic rate of 1 Mbps used to transmit the control packets and the physical header. Moreover, for each voice frame, a RTP/UDP/IP header has to be added. The proportion of this overhead is particularly high for small data packets.

As an example, the overhead budget for the transmission of a small packet of payload 80 bytes at 11Mbps is given in Tab.III. Note that the data part as well as the RTP/UDP/IP headers are sent at 11Mbps. The backoff has been set to 15 SlotTime.

TABLE III

OVERHEAD BUDGET FOR A PACKET WITH A PAYLOAD OF 80 BYTES, A BASIC RATE OF 1MBPS, AND A DATA RATE OF 11MBPS.

	Time transfer [μs]	Percentage
RTS (1Mbps)	352	21.4
CTS (1Mbps)	304	18.5
ACK (1Mbps)	304	18.5
PLCP Header (1Mbps)	192	11.7
3xSifs	30	1.8
Difs	50	3.0
Backoff	300	18.3
Data	58.2	3.5
MAC header	24.7	1.5
RTP/UDP/IP header	29.1	1.8
Total	1644.0	100

Since 1644.0 μs are needed for the transmission of 80 bytes, the user throughput is approximately 389 Kbps, which is in accordance with Fig.3. Now, a G711 call needs a bandwidth of 54,5Kbps if we take into account both directions and the voice activity. Thus, an upper bound for the capacity is $\lfloor 389/54.5 \rfloor = 7$. Simulation results show that for a voice quality requirement of $R = 70$, this value is reduced to 4.

B. Influence of P_{loss}

The influence of P_{loss} is now studied ($T_{nw} = 100ms$). On the one hand, increasing P_{loss} increases also the I_e parameter in the computation of R . On the other hand, the delay added by the dejittering buffer is reduced. Fig.8, 9, 10, and 11 illustrate this trade-off for resp. GSM-EFR, G711, G711 with PLC, and G723. Terminals are at 10 m of the AP. Similar results are also available for other distances.

With GSM-EFR, 1 voice call can be added to the previous capacity at 10 m, if $P_{loss} = 2\%$ is allowed on the dejittering buffer. Although G711 can take advantage of the PLC in term of voice quality, the maximum number of calls is still 4. Without PLC, G711 is very sensitive to packet losses: with $P_{loss} = 1\%$, the voice quality is already below our requirement. Let-

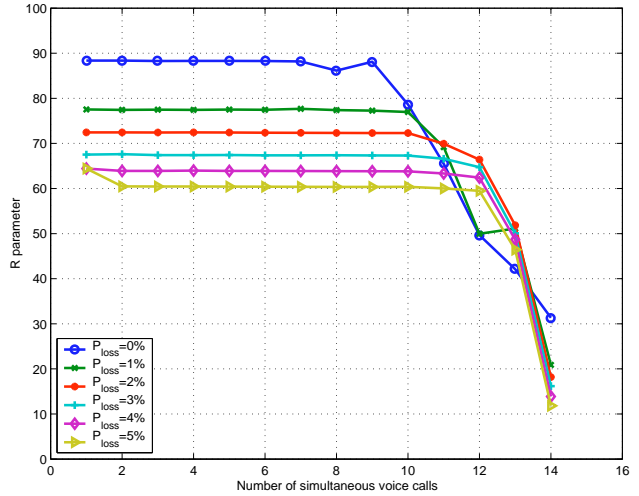


Fig. 8. R parameter vs. number of simultaneous voice calls, influence of P_{loss} , GSM-EFR.

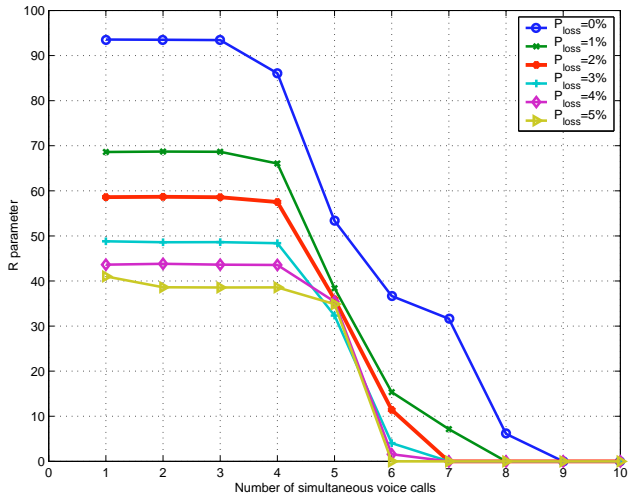


Fig. 9. R parameter vs. number of simultaneous voice calls, influence of P_{loss} , G711.

ting P_{loss} be 1% or 2% allows to have 18 G723.1 calls instead of 17.

C. Influence of the network delay

In this section, the influence of the network delay is studied by comparing the capacity with two extreme values $T_{nw} = 20ms$ and $T_{nw} = 100ms$.

The R parameter is given on Fig.12 for the two considered network delays and for terminals at 10m from the AP. The performance of G711 with PLC is not given because it is very similar to that of G711 when $P_{loss} = 0\%$. Reducing the network delay allows to obtain one more voice call for GSM EFR. There is however no gain for G711 and G723.

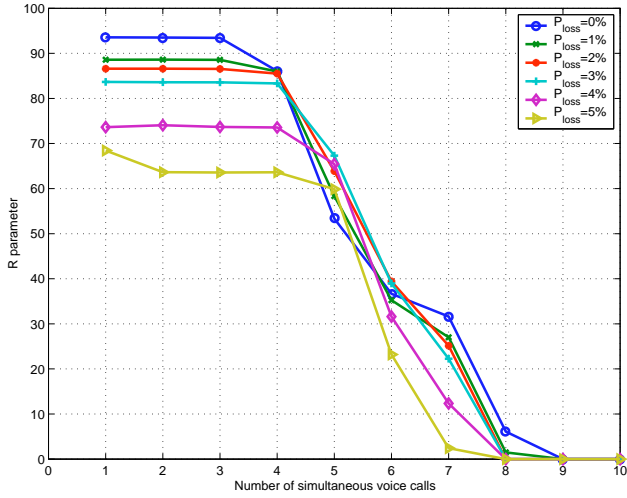


Fig. 10. R parameter vs. number of simultaneous voice calls, influence of P_{loss} , G711 with PLC.

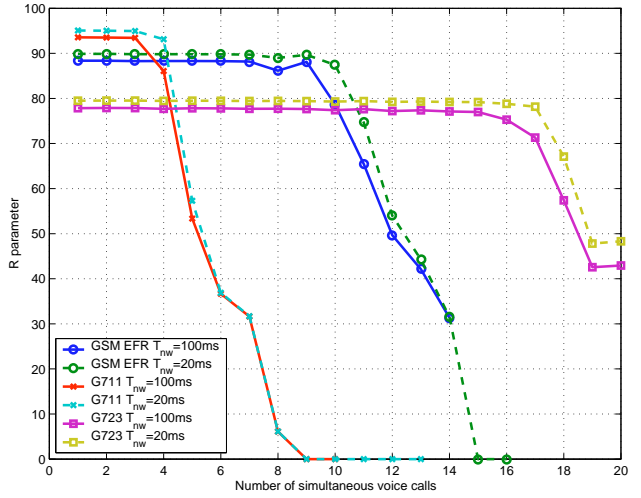


Fig. 12. R parameter vs. number of simultaneous voice calls at 10 m, influence of T_{nw} .

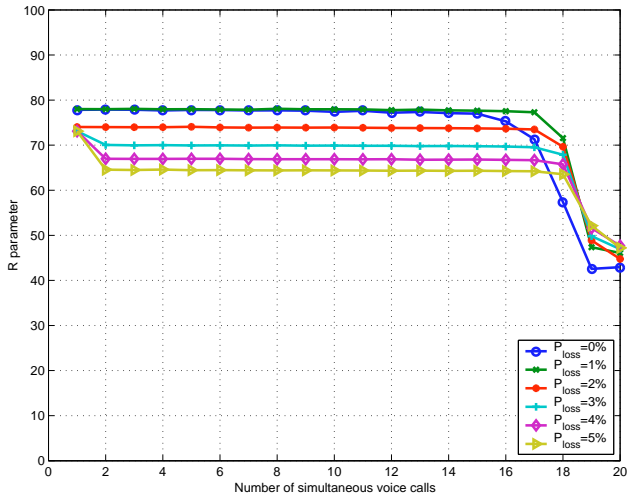


Fig. 11. R parameter vs. number of simultaneous voice calls, influence of P_{loss} , G723.1.

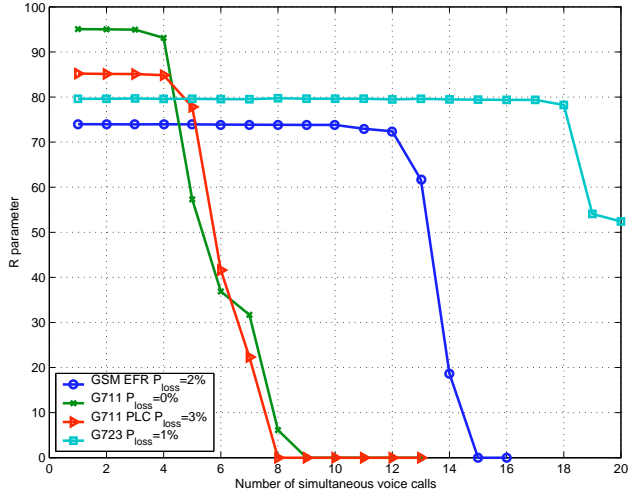


Fig. 13. R parameter vs. number of simultaneous voice calls at 10 m, $T_{nw} = 20ms$.

Fig.13 now shows the performance of the codecs with the best choice of P_{loss} among 0, 1, 2, 3, 4, and 5%. In this case, one more call can be obtained with G711 if PLC is used and $P_{loss} = 3\%$.

Thus, drastically reducing the network delay does not bring a huge increase in capacity. How can this phenomenon be explained? The explanation consists in the shape of the delay curves (Fig.14). For all codecs packet delays grow slowly with the number of voice calls. Even with $T_{nw} = 100ms$, the mouth-to-ear delay is very low, so that the voice quality is not impacted a lot by the network delay. As the number of voice calls reaches the capacity level, packet delays increase drastically: the voice quality is highly degraded. Hence, the network delay plays a role only

at the capacity limit.

V. CONCLUSION AND FURTHER WORK

In this paper, the VoIP capacity of IEEE 802.11b has been studied. Terminals are located at a fixed distance from the AP. An indoor propagation model has been used and a link adaptation mechanism has been taken into account. The influence of the codec, of the distance to AP, of the packet loss probability in the dejittering buffer, as well as of the network delay have been shown. Simulation results provide the following maximum achievable simultaneous voice calls: 5 for G711, 12 for GSM-EFR, and 18 for G723.1.

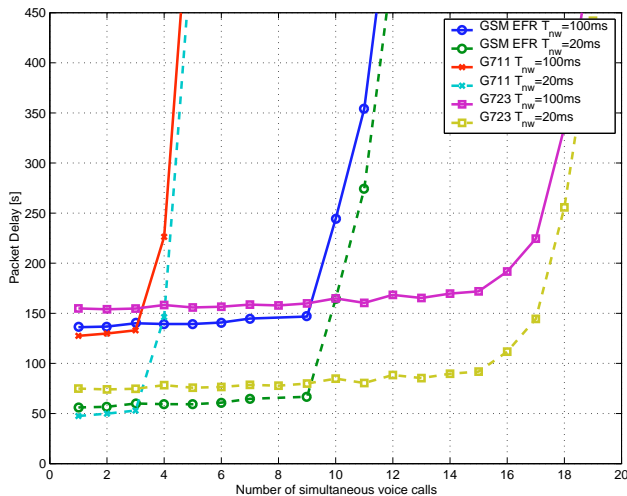


Fig. 14. Packet delay vs. number of simultaneous voice calls at 10 m, influence of T_{nw} .

Further work includes the fine-tuning of the MAC parameters. For example, the RTS/CTS handshake may not be needed on the downlink. Or the number of retransmissions could be reduced. Header compression could be also considered. At last, the concatenation of voice frames has to be investigated and could have a deciding role for reducing the MAC overhead.

APPENDIX

This section provides the E-model parameters for the simulations.

TABLE IV
E-MODEL PARAMETER'S VALUES.

SLR	8 dB
RLR	2 dB
LSTR	18 dB
STMR	15 dB
Ds	3 dB
Dr	3 dB
TELR	65 dB
WEPL	110 dB
qdu	1
Nc	-70 dBmOp
Nfor	-64 dBmp
Pr	35 dB(A)
Ps	35 dB(A)
A	5
Idte	0
Idd	0

Note that we have also $T = T_a = T_r/2$. Moreover, in VoIP, there is no hybrid echo. This is taken into account by the values of $TELR$ and $WEPL$.

ACKNOWLEDGEMENT

The authors would like to thank D. De Vleeschauwer for his help and his reports on the dejittering buffer mechanisms.

REFERENCES

- [1] S. Garg and M. Kappes, Can I add a VoIP call ?, Proc. of IEEE ICC'03, vol. 2, May 2003.
- [2] M. Veeraraghavan, N. Cocker, and T. Moors, Support of voice services in IEEE 802.11 wireless LANs, Proc. of IEEE INFOCOM'01, vol. 1, Apr. 2001.
- [3] J. L. Sobrinho and A. S. Krishnakumar, Distributed multiple access procedure to provide voice communications over IEEE 802.11 wireless networks, Proc. of IEEE GLOBECOM'96, vol. 3, Nov. 1996.
- [4] F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyanathan, A. Dutta, P. Agrawal, T. Kodama, and Y. Katsube, Voice performance in WLAN networks - An experimental study, Proc. of IEEE GLOBECOM'03, vol. 6, Dec. 2003.
- [5] A. R. Prasad, Performance comparison of voice over IEEE 802.11 schemes, Proc. of IEEE VTC'99, vol. 5, Sept. 1999.
- [6] M. G. Arranz, R. Agüero, L. Muñoz, and P. Mähönen, Behavior of UDP-based applications over IEEE 802.11 wireless networks, Proc. of IEEE PIMRC'01, vol. 2, Sept. 2001.
- [7] H.-H. Liu and J.-L. C. Wu, Packet telephony support for the IEEE 802.11 wireless LAN, IEEE Communications Letters, vol. 4, NO. 9, Sept. 2000.
- [8] S. Garg and M. Kappes, An experimental study of throughput for UDP and VoIP traffic in IEEE 802.11b networks, Proc. of IEEE WCNC'03, Mar. 2003.
- [9] N. Smavatkul, Y. Chen, and S. Emeott, Voice capacity evaluation of IEEE 802.11a with automatic rate selection, Proc. of IEEE GLOBECOM'03, vol. 1, Dec. 2003.
- [10] ITU-T Recommendation G.107, The E-model, a computational model for use in transmission planning, ITU-T, Mar. 2003.
- [11] ITU-T Recommendation G.113, Transmission impairments due to speech processing, ITU-T, Feb. 2001.
- [12] ITU-T Recommendation G.114, One way transmission time, ITU-T, May. 2003.
- [13] ITU-T Recommendation P.59, Artificial conversational speech, ITU-T, Mar. 1993.
- [14] D. De Vleeschauwer and J. Janssen, Voice performance over packet-based networks, Alcatel Technology White Paper.
- [15] J. Janssen, D. De Vleeschauwer, M. Büchli, and G. H. Petit, Assessing voice quality in packet-based telephony, IEEE Internet Computing, May/June 2002.
- [16] B. Goode, Voice over Internet Protocol, Proc. of the IEEE, vol. 90, NO. 9, Sept. 2002.
- [17] Texas Instruments, Voice over Packet, Texas Instruments White Paper, Jan. 1998.
- [18] J. Janssen and D. De Vleeschauwer, Designing voice over IP networks based on tolerable mouth-to-ear delay, Alcatel Corporate Research Center Report, TTD 454, Apr. 1999.
- [19] A. Van Moffaert and D. De Vleeschauwer, Adaptive dejittering mechanisms, Alcatel Network Strategy Group Report, TTD 461, Nov. 2000.
- [20] J. G. Proakis, Digital Communications, McGraw-Hill, 2000.
- [21] ns2.1b7, <http://www.isi.edu/nsnam/ns>,