

Voice Communications over Tandem Wireline IP and WLAN Connections*

Jerry D. Gibson¹, Bo Wei² and Sayantan Choudhury¹

¹Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106
gibson,sayantan@ece.ucsb.edu

²Huawei Technologies
1700 Alma Dr., Suite 100
Plano, TX 75075
wbo@huawei.com

Abstract- Voice communication over a tandem connection of a wireline packet switched network and an IEEE 802.11 access point is studied when there are packet losses in the wired IP network and bit errors in the WLAN link. Limited retransmissions are allowed, and packet loss concealment is utilized. We consider several alternative scenarios, including tandem free operation with G.711 in both links, G.711 over UDP in the wired link and G.711 with unequal error protection (UEP) in the WLAN, and G.729 in the wired backbone in tandem with G.711 UEP in the WLAN. To increase the number of calls supported by the WLAN access point, we also examine using AMR codecs at 12.2 kbps and at 5.9 kbps with forward error correction in the wireless link. The standardized PESQ MOS is used to evaluate delivered voice quality.

I. INTRODUCTION

As wireless local area networks (WLANs) such as IEEE 802.11a, b, g, and n proliferate and voice over Internet protocol (VoIP) becomes more widely accepted, voice calls over the tandem connection of a WLAN link and a wireline IP-based backbone will become common. Initially it may seem that this is a straightforward packet voice communications connection and that there is little to be investigated. However, the unreliable transmission channel of the wireless LAN link and the CSMA/CA access protocol used in the WLAN can cause significant packet losses or high latency if the voice application is not properly integrated into the WLAN operation. Furthermore, because of the possibility of packet losses in both the wireless and wireline links, packet loss concealment (PLC) can become important. Additionally, the WLAN user and the wireline backbone may support/employ different voice codecs, which can require transcoding at the network interface.

In this paper, we consider a system where a VoIP wireline backbone is connected in tandem with a VoWLAN link via an access point as shown in Fig. 1. In this tandem system, packet losses can occur in the wireline backbone network while contention and channel noise can result in bit errors and/or packet losses in the wireless LAN channel.

We investigate the following scenarios:

- Tandem free operation (TFO) with G.711 in both links
- A tandem connection of packetized G.711 over UDP in the wireline backbone and G.711 with unequal error

protection (UEP) in the WLAN link with PLC at the network interface

- A tandem connection of packetized G.729 in the backbone and G.711 with unequal error protection (UEP) in the WLAN link with PLC at the network interface

We also investigate the use of a lower rate voice codec in the WLAN link. In particular, we investigate:

- Using the Adaptive Multi-Rate (AMR) voice codec at 12.2 kbps with unequal error detection (UED) in the WLAN link in tandem with G.711 or G.729 in the wireline backbone
- Using the AMR codec at 5.9 kbps with forward error correction (FEC) over the WLAN link in tandem with G.711 or G.729 in the wireline backbone
- Switching between AMR at 12.2 kbps with UED and AMR at 5.9 kbps with FEC, depending upon WLAN link conditions, in tandem with G.711 or G.729 in the wireline backbone

Comparisons of the maximum number of (two-way) voice calls supported on an IEEE 802.11a access point for the several transmitted data rates are given to complete the comparison. The standardized PESQ MOS is used throughout to evaluate delivered voice quality.

II. ISSUES IN TANDEM WIRELINE VOIP/VOWLAN CONNECTIONS

G.711 and G.729 are two standardized voice codecs that are widely used in VoIP systems and hence these codecs are used as the wireline backbone codecs in this work [1, 2]. We use the standard error concealment scheme, ITU G.711 Appendix I [3] for G.711, and the recommended error concealment scheme for G.729 [2]. Transcoding is another main source of distortion for tandem voice transmission [4], and we consider scenarios both with and without transcoding between codecs at the network interconnections. We take advantage of possible cross layer designs by jointly considering several layers of the protocol stack spanning from application layer parameters to physical transmission. In the G.114 ITU-T Recommendation [5], the maximum delay for “most users satisfied” category is 280 ms. To be conservative, we set the

* 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.

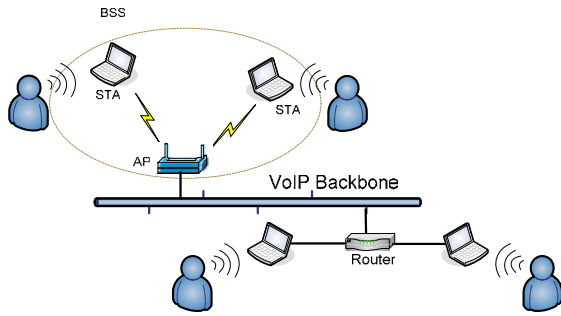


Figure 1. VoWLAN/VoIP system

end-to-end delay budget to 200 ms, and choose the packet drop timer, which is the amount of time we wait for a packet, for voice traffic in the WLAN as 40 ms.

Despite the efforts on improving the robustness of the physical layer in wireless channels, some of the errors propagate to the link layer. If the errors are detected by the link layer frame check sequence, each MAC frame with one or more errors is discarded by the link layer regardless of the number or location of the errors. This classical layered stack may not be suitable for multimedia transmissions because even partially damaged packets may lead to at least a coarse or noisy version of the decoded multimedia frame, which may be more useful than a discarded frame. Therefore, tradeoffs can be made between the bandwidth saved by reducing retransmissions and the quality lost due to the bit errors allowed in source decoding.

The flexibility of UDP is increased by providing a partial checksum that optionally covers only the UDP/IP headers or headers plus the “most important” payload bits [6]. “MAC Lite” [7] is proposed to address this problem in WLANs where the MAC layer abandons retransmissions and passes partially corrupted packets to higher layers. The MAC layer can be further modified to protect only the MAC header, the entire header, or the entire header along with part of the “important” payload [8], and thus provides more flexibility in cross-layer design for multimedia traffic.

III. G.711/G.729 WIRELINE VOIP IN TANDEM WITH G.711 OVER WLAN

A. Cross-layer Design

The original IEEE 802.11 MAC layer has a CRC that is calculated over the entire packet, including all headers and the voice data portion, and in the event of bit errors in this area, the MAC will drop the packet and ask for retransmissions until a maximum retransmission limit or drop timer threshold is reached. In our work, in order to improve the WLAN link efficiency and limit retransmissions, we allow bit errors in the voice data portion. A cross-layer design must be performed to accomplish this task. Note that because of the non-uniform bit sensitivity of any coded speech streams, we further classify bits into two or more groups with different perceptual importance.

In our UEP scheme, the MAC CRC only protects headers and part or all of the voice payload is protected using forward

error control coding and checked with an additional higher layer CRC. The packet dropping procedure is: If the MAC layer CRC check fails, the packet is dropped and retransmission is requested, otherwise, the packet is retained. For the retained packets, after the channel decoding process for the most important bits, we go through the application layer CRC check for these bits. If this CRC fails, error concealment is applied; otherwise, source decoding is performed even if there are errors in less important bits. Note that at the transport layer, the UDP checksum has to be disabled so that the partially corrupted speech frames can pass through UDP and go up to the application layer.

B. Tandem Simulations

We consider the scenario where two VoWLAN users communicate with two VoIP users via the same access point and IP backbone as shown in Figure 1. Two VoWLAN users are simulated to introduce some contention, without saturating the links. In addition to the packet losses caused by contention and delay, non-ideal WLAN channels are simulated where bit errors can occur during transmission over the air link.

In our simulations, the IEEE 802.11b protocol with the distributed coordination function (DCF) MAC layer is assumed and ns2 [9] is used as the system simulator. We use the RTP/UDP/IP transport layer protocol in the VoIP backbone, while a modified version that allows bit errors, is used for the WLAN portion on top of the 802.11b MAC/PHY protocol. In this way, an all-IP network is set up. Robust header compression (ROHC) is assumed to compress the UDP/IP header [10] to an average header length of 4 bytes. Since MAC header compression is still under investigation, we consider the standard 28 bytes MAC header length in this work.

In the VoIP backbone, the two most popular VoIP codecs, G.711 at 64 kbps and G.729 at 8 kbps are employed. For the VoWLAN link, in addition to investigating G.711, we propose an unequal error protected version of G.711 (referred to here as G.711UEP). In G.711UEP, we group the Most Significant Bits (MSBs) and Least Significant Bits (LSBs) in the G.711 coded stream separately. MSB bits are then protected with a rate $\frac{1}{2}$ convolutional code. This results in a 120 byte payload for a 10 ms speech frame (80 byte payload + 40 bytes for FEC of MSBs). Since the proposed MAC CRC does not check for bit errors that occur in the “voice data” portion, only bit errors in the headers result in retransmissions and possibly packet losses. PLC is performed when packets are discarded due to the drop timer.

For the WLAN link, a 40 ms drop timer is assumed in the simulations. PESQ [11] is used to provide objective MOS values for each test scenario and various VoIP link packet loss rates and VoWLAN bit error rates.

C. Reference System (G.711TFO)

For a reference system, G.711 is used as the speech codec for both the VoIP and WLAN channels. Error concealment is performed at the end user node but not in the intermediate

node, which is the wireless access point in this case. We refer to this reference system as “G.711TFO”.

Note that retransmissions are not allowed in the backbone since the UDP protocol is used. With our TFO configuration, in the WLAN channel, a MAC layer packet is dropped whenever bit errors are detected in the headers including MAC/RTP/UDP/IP headers; otherwise, the packet is kept and decoded regardless of any possible bit errors in the voice data portion.

D. G.711 Synchronous Transcoding with Unequal Error Protection (G.711ST-UEP)

We also consider a tandem system for the VoIP/VoWLAN system where G.711UEP is used for the WLAN channel for bit error robustness and G.711 is used in the backbone. We have intermediate noisy channel compensation here that is accomplished by error concealment schemes at the network interface. We refer to this new system as “G.711 Synchronous Transcoding with Unequal Error Protection”, or “G.711ST-UEP”. We set the PLR range of our simulation as $2 \times 10^{-3} \leq PLR \leq 2 \times 10^{-1}$, and the packet drop timer is set as 40 ms in the WLAN link.

For G.711UEP in the WLAN channel, the bit errors remaining in the MSBs after convolutional decoding result in retransmissions while leaving any bit errors in the LSBs unattended. Thus, although we show the WLAN channel condition variation based on bit error rate, the actual losses seen from the application layer point of view consist of a mixed packet loss/bit error scenario. Note that G.711UEP has 120 byte packets which are longer than the pure G.711 coded packets (80 bytes) because of the extra protection to the MSBs. Simulation results of the G.711TFO and G.711ST-UEP systems are given in Figure 2 with different VoIP backbone PLRs and WLAN channel BERs. The plots are based upon 8 speech sentences of a total 24 seconds duration and 200 realizations for each pair of channel conditions.

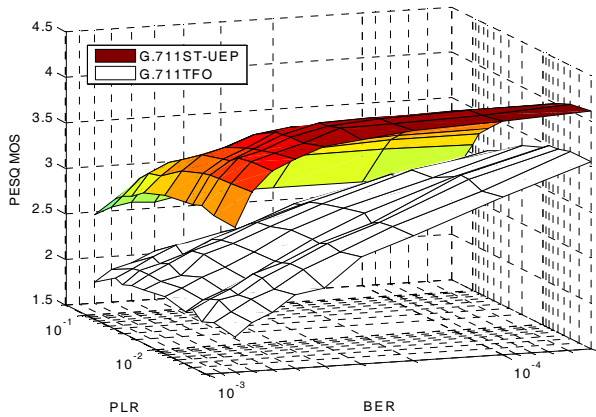


Figure 2 Speech Quality comparison between proposed G.711ST-UEP and G.711TFO system (40 ms drop timer in WLAN)

In Figure 2, we see that the TFO system is very sensitive to both packet losses and bit errors. Even when the BER is as low as 5×10^{-5} and with good backbone channel conditions

($PLR = 2 \times 10^{-3}$), the delivered speech quality is still around 3.6 in MOS compared to the original G.711 coded speech quality without channel errors at about 4.1. When both channels are poor ($PLR = 2 \times 10^{-1}$), the speech MOS value drops to 1.7. With the extra protection for the MSBs in the G.711ST-UEP system, bit error robustness is enhanced so that the overall tandem speech quality is improved to 2.5 when both channels are poor and to more than 4.0 when the channels are good.

An improvement of 0.4-1.5 in MOS is observed across the entire set of simulated channel conditions. Notably, the packet length of the G.711ST-UEP system is 0.36 times longer, and as a result, the packet error probability is higher for the G.711ST-UEP system. Despite the increased packet loss rate, the system is very effective compared to TFO.

E. G.729 Asynchronous Transcoding with Unequal Error Protection (G.729AT-UEP)

When G.729 is used in the backbone, asynchronous transcoding must be performed between G.729 in the VoIP backbone and G.711 which is in the WLAN link. We refer to this new system as “G.729 Asynchronous Transcoding System with G.711UEP” (G.729AT-UEP). The setup in the WLAN link is the same as that in Sec. III.C.

Simulation results are given in Figure 3. We notice that the MOS value drops by around 0.6 by using G.729 in the backbone, because the inherent MOS of G.729 is less than that of G.711 and because of the poor G.729 error concealment performance and error propagation problem.

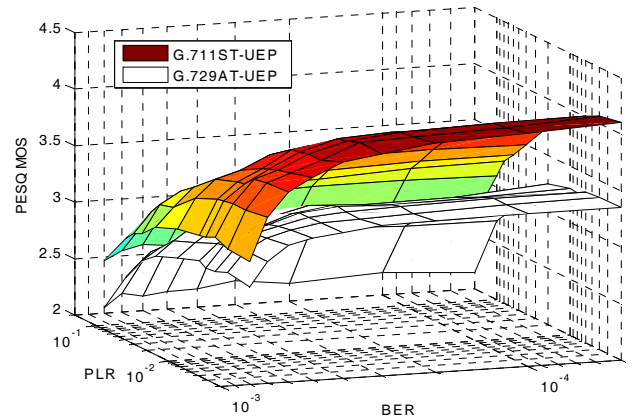


Figure 3 Speech Quality comparison between proposed G.711ST-UEP and G.729AT-UEP system (40 ms drop timer in WLAN)

IV. G.711 WIRELINE VOIP IN TANDEM WITH AMR CODECS OVER WLAN

In order to increase the number of supported voice calls and to maintain performance over a wider range of WLAN channel conditions, we now investigate the use of the GSM Adaptive Multi-Rate (AMR) codecs at 5.9 kbps and 12.2 kbps over the WLAN in tandem with G.711 and G.729 in the wireline backbone.

A. Unequal Error Detection with Cross-layer Design

We explore Unequal Error Detection (UED) for a MR12.2 coded stream, and introduce link adaptation across the application layer. UED does not need extra forward error protection bits and results in a shorter packet size. However, a cross-layer design still needs to be applied. With UED, we only let the MAC CRC check the entire header along with the most important bits. Whenever bit errors are detected in this area, the MAC frame is dropped in the MAC layer and retransmission is required. Otherwise, this frame is retained and transferred to the higher layer.

We propose an unequal error detected NBAMR [12] codec at rate 12.2 kbps (we refer to this setup as MR12.2UED) working in the VoWLAN system. MR12.2 encodes 20 ms speech frames with 244 bits, where 81 bits are classified as Class A bits, which are the subjectively most significant bits. Also, we utilize the ROHC compressed UDP/IP header with an average length of 4 bytes. In this UED scheme, only 43 bytes of header plus the 81 Class A bits are protected by a CRC, and the 21 Class B bytes are left unprotected.

We also investigate the MR 5.9 kbps codec. The entire MR5.9 coded speech frame bits are forward error protected with a rate $\frac{1}{2}$ convolutional channel code (we refer to this setup as “MR5.9FEC”). In the MR5.9FEC scheme, the MAC CRC only checks the headers. Packets are retained and passed to higher layers if the MAC CRC check on the headers passes. With the rate $\frac{1}{2}$ convolutional channel encoding and tail bits, the MR5.9FEC payload size is 32 bytes for each 20 ms packet, which is the same as that of the MR12.2UED packet. 4 bytes ROHC compressed UDP/IP header and 28 bytes MAC header are also used.

B. VoWLAN Tandem System with Source/Channel Rate Reallocation

With the different cross layer design and protection/detection schemes as described before, MR12.2UED and MR5.9FEC have exactly the same MAC layer packet size (64 bytes) for a 20 ms speech frame. This packet size is only about 42% of the G.711UEP packet length.

Tandem System Setup

In the VoWLAN/VoIP system, G.711 is investigated as the VoIP backbone speech codec. The RTP/UDP/IP transport layer protocol is applied for the VoIP backbone, while a modified version (with UDP checksum disabled) that allows bit errors, is used for the WLAN portion on top of the 802.11b MAC/PHY protocol. Whenever a packet is lost in the backbone wireline network, no retransmission is allowed and error concealment is performed. In the WLAN, however, MAC layer retransmissions are allowed based on our modified MAC CRC until it reaches the drop timer value of 40 ms.

For the forward link, error concealment is performed at the access point and the resulting speech streams are re-encoded with either the MR12.2 or MR5.9 codec and transmitted over the WLAN channel.

C. Adaptive Source/Channel Rate Reallocation

Adaptive reallocation of the available data rate between source and channel coding has been applied previously in the mobile system to achieve better speech quality under certain channel conditions [13]. Switching among several source/channel rate sets can be performed using channel estimation and preselected thresholds. We evaluate the performance of an adaptive system that switches between MR5.9FEC and MR12.2UED in the WLAN.

D. Tandem Simulations

In the VoIP backbone, G.711 is used while the MR5.9FEC and MR12.2UED schemes we discussed above are used in the WLAN link. For the WLAN link, a 40 ms drop timer value is assumed.

G.711 with Adaptive Asynchronous Transcoding

In Figure 4, the forward tandem system performance of G.711 in the IP backbone with MR12.2UED or MR5.9FEC in the WLAN link is shown. Due to the good asynchronous transcoding properties of G.711, speech quality is only slightly degraded by transcoding with MR12.2 or MR5.9. However, as expected, when both channels are good (roughly when $BER < 4 \times 10^{-4}$ and $PLR < 10^{-1}$), the tandem system with MR12.2UED outperforms the one with MR5.9FEC. When both channels are degraded, the system with MR5.9FEC provides much better performance due to the forward error protection. The transition line between these two systems is at about $BER = 4 \times 10^{-4}$ (see the vertical crosscut in Figure 4).

If we have the flexibility of reallocating the source and channel bit rate by performing adaptation between MR5.9FEC and MR12.2UED, we can achieve performance along the envelope of the two surfaces in Figure 4. We call this new adaptation system “G.711 with Adaptive Asynchronous Transcoding” (G.711AAT). Performance of the G.711AAT system is shown in Figure 5 and compared with the G.711TFO and G.711ST-UEP systems discussed in III.C and III.D.

From Figure 5, we see that much better overall performance in terms of MOS value is obtained by the G.711AAT system compared to the reference G.711TFO system. In particular, G.711AAT can provide quality that is competitive with the original G.711 coded speech when both channels are good and can also provide acceptable quality ($MOS \geq 3.0$) over a wider range of channel conditions, i.e. $BER < 6 \times 10^{-4}$ and $PLR < 15\%$, (see the horizontal crosscut in Figure 4). Although G.711ST-UEP can extend the system operational range until $BER < 7 \times 10^{-4}$, it consumes 1.4 times more bandwidth than G.711AAT, which may not be desirable for system capacity considerations.

V. WLAN VOICE CAPACITY

Comparisons of the maximum number of (two-way) voice calls supported on an IEEE 802.11a access point for the several transmitted data rates are given in Table I. A few interesting conclusions are evident. First, G.711 with UEP does not incur much of a penalty in the number of voice calls

VI. CONCLUSIONS

It is clear that although transcoding can reduce quality, PLC at the network interface can be useful. UEP and UED of MSBs can significantly improve performance for all codecs, and retransmissions can be reduced with PLC and cross layer designs. Switching between codec/error control methods in the WLAN improves performance, extends the operational region, and maximizes WLAN call capacity for the conditions examined. The broad lesson is that modifications of the MAC to allow packets with bit errors can improve performance and reduce retransmissions.

REFERENCES

- [1] ITU-T Recommendation G.711, Pulse code modulation (PCM) of voice frequencies, ITU-T, November 1988.
- [2] TU-T Recommendation G.729, Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP), ITU-T, March 1996.
- [3] TU-T, G.711, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711, Sept. 1999.
- [4] J. D. Gibson, "Speech Coding Methods, Standards, and Applications," *IEEE Circuits and Systems Magazine*, pp. 30-49, Fourth Quarter 2005.
- [5] ITU-T Recommendation G.114, One Way Transmission Time, ITU-T, May 2000.
- [6] L-A. Larzon, M. Degermark, S. Pink, L-E. Jonsson, and G. Fairhurst, The UDP-lite Protocol, draft-ietfsvwg-udp-lite-02.txt, August 2003.
- [7] S. A. Khayam, S. S. Karande, M. Krappel, and H. Radha, "Cross-layer Protocol Design for Real-Time Multimedia Applications over 802.11b Networks," *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, July 2003.
- [8] I. Chakeres, H. Dong, E. M. Belding-Royer, A. Gersho, and J. D. Gibson, "Allowing Errors in Speech over Wireless LANs," *Proceedings of the 4th Workshop on Applications and Services in Wireless Networks (ASWN)*, August 2004.
- [9] UCB/LBNL/VINT, "Network Simulator - ns - version 2," URL: www.isi.edu/nsnam/ns, 1997.
- [10] C. Bormann, C. Burmeister, M. Degermark, H. Fukushima, H. Hannu, L-E. Jonsson, R. Hakenberg, T. Koren, K K. Le, Z Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, RObust Header Compression (ROHC): Framework and four Profiles: RTP, UDP ESP and uncompressed, RFC 3095, July 2001.
- [11] ITU-T Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs, ITU-T, 2001.
- [12] 3G TS 26.090 "Adaptive Multi-Rate (AMR) Speech Transcoding". 1998.
- [13] B. Wei, H. Dong, and J. D. Gibson, "Application of NB/WB AMR Speech Codecs in the 30kHz TDMA System," *IEEE Transactions on Wireless Communications*, pp. 1897-1902, November 2004.

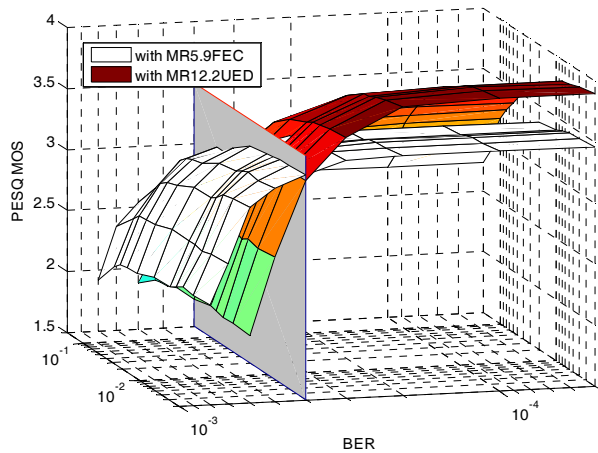


Figure 4 Tandem performance comparison between system with MR12.2UED and the one with MR5.9FEC (with G.711 in backbone)

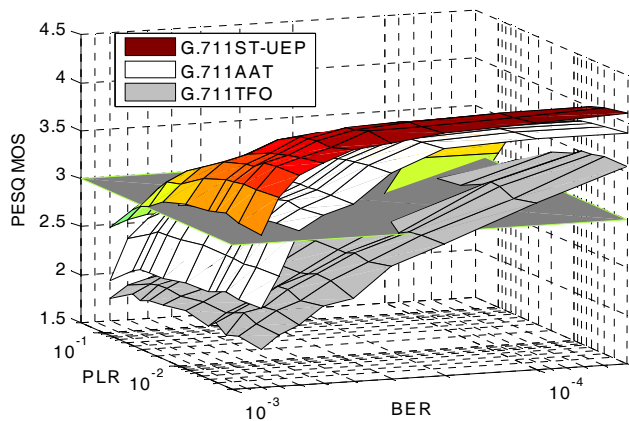


Figure 5 Performance comparison among G.711AAT, G.711ST-UEP and G.711TFO

Table I. Maximum Number of Voice Calls Supported for Different Codecs and Payload Sizes over IEEE 802.11a WLANs

Data Rate (Mbps)	6	9	12	18	24	36	48	54
/Codec								
10 ms G.711	16	19	23	26	29	31	32	32
10 ms G.711 with UEP	14	17	20	24	27	29	31	32
20 ms 12.2 Kbps AMR with UED	42	48	55	59	64	67	69	69
20 ms 5.9 Kbps AMR	46	52	58	62	65	69	69	71
20 ms 5.9 Kbps AMR with FEC	41	48	54	59	64	67	69	69

supported compared to G.711 without any error protection. Second, the number of voice calls supported can be more than doubled by using the AMR codecs. Third, if a minimum MOS value of 3.5 is specified, the AMR 5.9 codec cannot be used at all.