CROSS-LAYER SOLUTIONS TO PERFORMANCE PROBLEMS IN VOIP OVER WLANS

F. Maguolo, F. De Pellegrini †, A. Zanella and M. Zorzi

Department of Information Engineering, University of Padova,
@-mail:{maguolof,depe,zanella,zorzi}@dei.unipd.it
via Gradenigo 6/B, 35131 Padova, Italy.
† also CREATE-NET
via Solteri 38, 38100 Trento, Italy
@-mail:fdepellegrini@create-net.it

ABSTRACT

The design of WLANs was meant to extend Ethernet LANs in the most transparent way, but no particular mechanism was deployed in order to support real-time applications natively. At present VoIP calls are becoming customary, and IEEE802.11 WLANs must face the provision of guaranteed quality of service. In practice, QoS should be provided somehow a posteriori on top of the existing standard. In this paper, we address some concerns on the efficiency of WLANs for VoIP provision already remarked in literature and analyze possible solutions to increase the voice capacity of DCF IEEE802.11 WLANs. We consider two candidate solutions, the VA [1] and the M-M [2] cross-layer schemes. The efficiency of such mechanisms is evaluated in order to assess the performance gain compared to existing solutions. We provide extensive simulation results, proving that the advantage is significant, while requiring minor changes compared to the current IEEE802.11 standard.

I. INTRODUCTION

The leading RF solution for wireless indoor connectivity is represented by the IEEE802.11 standard [3]. Recently, increasing attention has been devoted to the provision of VoIP support over IEEE802.11 WLANs, and the reason for the increasing interest in such direction is twofold. First, there exists a clear economical advantage for mobile phone service providers, since leveraging the free ISM band, in principle, it is possible to relieve part of the indoor voice traffic from cellular networks. In fact, as mobile phones start beeing equipped with IEEE802.11 RF interfaces and suitable software, handing off users from a cell into existing indoor hot-spots represents an appealing economic advantage for providers. In such scenario, the user could leverage on the WLAN presence (possibly at lower fares) and the provider could free some channels for additional allocation.

The second reason comes from the spreading habit of LAN users to rely on VoIP telephony [4]: with the ongoing migration to wireless connectivity, such users will represent a source of VoIP traffic over WLANs. Some of such applications, anyway, are usually designed on the wired LAN scenario, typically Ethernet LANS. In the wired scenario, the MAC and the PHY layer are different, since the MAC is regulated through the CSMA/CD technique and attenuation or interference effects are much milder than in the case of radio propagation. Thus, even though the nominal rate over a modern IEEE802.11 WLAN link is close to the rate offered by a 100 Mb/s Ethernet link, several authors [2, 5–7] claim that the current implementation of VoIP over IEEE802.11 WLANs carries some structural inefficiencies and, in the end, at present is not satisfactory.

A further reason why the problem of VoIP over WLANs deserves investigation, is related to fairness issues. Voice applications, in fact, if possible (basically when the network management permits) adopt UDP transport, trading off flow control protection for faster delivery. This is, for example, the case of Skype. Such a behavior increases the injection of non TCP-friendly flows in the Network. UDP flows, in turn, compete unfairly with TCP traffic, so that, when such flows are served inefficiently, unfairness towards TCP regulated traffic may experience a significant degradation.

The paper is organized as follows. In Section II we describe the network scenario and the major problems of the existing customary configuration. In section III we recall the results found by authors of related papers. In Section IV we describe the novel VA cross-layer scheme, and the reference M-M scheme. In section V we describe the performance figures and compare the two cross-layer schemes under analysis. The last section provides some concluding remarks.

II. NETWORK SCENARIO

The network considered in this work is a WLAN where some stations establish VoIP sessions with terminals outside the WLAN. The Access Point (AP) forwards the voice traffic towards a gateway which could connect either to a PSTN network or to the Internet. The WLAN is a set of IEEE802.11 terminals which communicate through an AP, and each located within radio range: thus, the terminals and the AP, form a one-hop network [8] under the Infrastructure mode. As introduced before, this is the typical configuration, which is faced by almost every VoIP over WLAN implementation. In this case, no direct communication between terminals is not permitted, and all packets are destined to the AP first, and then forwarded towards the destination. Terminals, and the AP as well, employ the Distributed Coordination Function access mechanism. We recall that, according to the standard [3], there exist an alternative to the DCF mode, namely the Point Coordination Function (PCF), and (not surprisingly) according to some authors PCF would provide a more performing solution for VoIP support [9, 10]. Nevertheless, the DCF configuration described above is practically mandatory, since a sudden replacement of millions of existing WLANs cards and APs is not likely to occur soon.

In the use of DCF, problems such as exposed node, hidden node and related issues might represent a threat for correct operation of WLANs. In the one-hop scenario considered here they are not taken into account, but, in general, this assumption is not reasonable, since the RF fields cannot be easily confined and near-far effects do occur anyway. Nevertheless, during the APs placement, a careful frequency assignment can limit such effects. In particular, this is the case when the frequencies of channels assigned to each access point do not overlap with those assigned to neighboring APs. To this aim, the IEEE802.11 standard [3] makes available 12 channels to be assigned to WLANs.

As pointed out in [6, 11, 12], the DCF regulated Infrastructure mode has structural limitations. One is the bottleneck at the AP, since the long term fairness provided by IEEE802.11 makes the share reserved to the AP equal to the one reserved to terminals under an incoming rate n times larger. Also, in a WLAN, the perceived quality at the receiver side is required to be rather insensitive to the presence of alien traffic, because several IEEE802.11 terminals may establish concurrent data flows towards the Internet and use the AP connectivity for data trasfer [13].

In this paper we show that non-disruptive cross-layer solutions enhance the performance of VoIP over the DCF Infrastructured configuration.

III. RELATED WORK

Several papers pointed out concerns on the use of the DCF regulated Infrastructure mode. In particular, a quite common feature of voice codecs is the small payload size of the packets. In the case of the ITU G.711 codec, which packetizes a plain PCM flow, the typical payload is of 80 bytes per packet. Enhanced codecs have even smaller payload. Thus, several authors found that the overhead added by the protocol layers in order to convey voice packets leads per-se to a very small number of supported voice sessions. A voice packet is added 12 bytes of RTP overhead, 8 bytes of UDP header, 20 bytes of IP header, and, finally, 34 bytes of MAC header and checksum, thus adding 74 bytes of overhead.

The works in [2, 5, 11] determined the performance of an IEEE802.11b WLAN using DCF with Infrastructure mode in terms of the number of sustainable VoIP flows, for several codecs. The analytical study and related simulations show that only 6 connections are possible using G.711 codec (64 kbps) and 7 using G.729 codec (8 kbps). When compared to the nominal bitrate, i.e. 11Mb/s, this is quite surprising, since the bitrate of voice sessions is 2 orders of magnitude smaller [2, 5].

As remarked in [5], the plain reduction of the protocol overhead, which can be beneficial in the wired case, in the case under consideration is rather limited, since a major overhead is introduced by the exponential backoff mechanism. For example, in the case of the *IEEE 802.11g PHY*, with 54 Mbit/s both for data rate and basic rate, and a G.711 codec, we obtain for the transmission time of a voice packet $T = T_{voice} + SIFS + T_{ACK} + DIFS = 118\mu s$, whereas the most optimistic estimation of the time spent for the average backoff stage is 67μs (we did not account for collisions or for the backoff freezing proces). Thus, due to the backoff procedure, decreasing the size of voice packets does not represent an advantage in this configuration. All the problems described above are exacerbated by the need for coexistence of voice traffic with alien traffic [2].

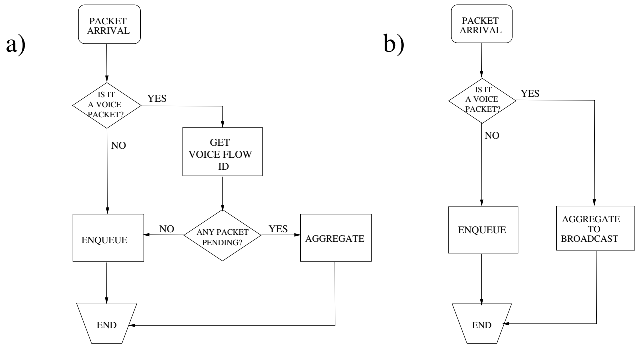

Fig. 1. a) VA scheme flowchart. b) M-M scheme flowchart

Despite the remarks above, few solutions have been proposed to increase the voice capacity of WLANs, defined as the maximum number of sustainable VoIP calls for a given codec. A solution employing protection of VoIP through a preemptive queue see [13]. A solution, based on a booster mechanism, is proposed in [10].

We will describe the solution appeared in [2]. The authors propose a scheme which enhances the performance of VoIP over a WLAN with no need for a massive replacement of currently deployed terminals, and modifying the AP only. Thus, the work in [2] relieves the scalability problems of VoIP over IEEE802.11 efficiently and being largely compatible with existing terminals. Our cross-layer technique, the VA scheme, represents a complementary solution. The main contribution of this paper is a thorough performance evaluation and a direct comparison of the two schemes.

IV. CROSS-LAYER SOLUTIONS

Vertical Aggregation: The cross-layer scheme we propose is named *Vertical Aggregation* (VA) since it works along the same flow. The main advantage of our solution is that it enhances voice capacity using a plain IEEE802.11 MAC protocol, and adopting an additional application-aware module, logically placed above the MAC layer. Such a module monitors all the active VoIP flows and take actions accordingly: since we discriminate among different applications at the link layer, we purposely break the transparency rule. Such a mechanism was inspired by the findings of [7]: the increase of the voice codec interframe generation interval increases voice capacity. The rationale is that, when the number of VoIP flows increases, as the service time at the MAC layer becomes larger, several packets of the same VoIP flow are queued waiting for service. We compound all pending packets with an aggregation procedure and lower the packet departure rate. Forming a larger packet size, we counteract to the service time increase. In Fig. 1a), we described the behavior of the aggregation module through a flowchart. Packets from the IP layer to the MAC layer are classified as either VoIP packets or data packets. Non-voice packets are simply enqueued, whereas VoIP packets are inspected IP and RTP headers to recognize the voice flow they belong to [2]. Hence, incoming VoIP packets are compounded together with VoIP packets belonging to the same flow, and waiting for MAC service. The algorithm works both the AP and at terminals. Terminals will typically perform the aggregation of packets over a single VoIP

TABLE I

PARAMETERS ADOPTED FOR THE IEEE802.11G NETWORK.

Parameter	*IEEE 802.11b*	*IEEE 802.11g*
RTP layer overhead	12 bytes	
UDP layer overhead	8 bytes	
IP layer overhead	20 bytes	
MAC layer overhead	34 bytes	
ACK packet size	14 bytes	
SlotTime (T_{Slot})	$20\mu s$	$9\mu s$
SIFS	$10\mu s$	$10\mu s$
$DIFS(SIFS + 2 \cdot T_{Slot})$	$50\mu s$	$28\mu s$
Preamble length ($T_{PLCPpreamble}$)	$144\mu s$	$16\mu s$
PLCPHeaderLength ($T_{PLCPheader}$)	$48\mu s$	$4.296\mu s$
SIGNAL length (T_{PLCP_SIG})	$8\mu s$	$4\mu s$
ShortRetryLimit	7	7
Signal extension (T_{SE})	N/A	$6\mu s$
CW_{min} (units of SlotTime)	31	15
CW_{max} (units of SlotTime)	1023	1023

flow, whereas the AP needs to apply the same aggregation procedure to several multiplexed flows.

M-M scheme: The scheme proposed in [2], relieves the bottleneck formed at the AP using a multiplex-multicast scheme (M-M): pending packets at the AP can be served simultaneously, aggregating them into a large broadcast frame, every T seconds. The M-M scheme prioritizes the AP through a novel interframe spacing interval, the Multicast Interframe Spacing (MIFS): this is due since broadcast over IEEE802.11 is unacknowledged and collisions might impair the efficiency of such a scheme. Using MIFS, the AP does not collide with terminals since it gains medium access before any other station and transmits immediately. Also, the bitrate used for broadcast frames is assumed to range over the possible set of data rates.

The M-M scheme is reported through a flowchart in Fig 1b): as in [2], the inter-frame generation at the AP, i.e. parameter T, is a free parameter. For an homogeneous scenario, a choice indicated in [2] is to set T as the voice packet generation interval.

V. SIMULATION RESULTS

In this section we report the performance figures of the solutions described before. The target is to support as many voice sessions as possible per AP [5]. The simulation results were obtained using the NS2 simulator [14], with the parameters of Tab I. In these series of simulations, we considered constant bit rate voice codecs because this is a worst case compared to the use of variable bit rate codecs, due to the larger average throughput required. Measuring the WLAN voice capacity in this case has practical interest: CBR codecs are quite popular due to their simplicity. The G.711 codec, in particular, is adopted by Microsoft MSN Messenger for the VoIP application [13]. Also, the use of a constant bit rate encoding with no silence suppression apparently is a relevant commodity feature in maintaining UDP bindings and filling persistently the TCP pipe [4].

First, we measured the average delay experienced by voice packets at the increase of the number of the voice stations and obtain a first general description of the impact of voice stations on the stability of the WLAN. Second, we measured the useful ratio, is the ratio of the number of packets which arrived in time and the total number of transmitted packets: it is a measure the perceived speech quality. For a G.711 codec, tests on commercial devices indicate that 5% loss rate still provide a fair decoded speech quality. We resorted to ITU recommendations, and assumed

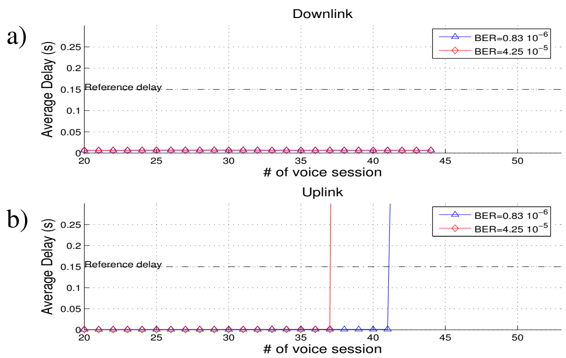

Fig. 2. M-M scheme. Average voice packet delay vs. number of voice sessions; a) Uplink b) Downlink.

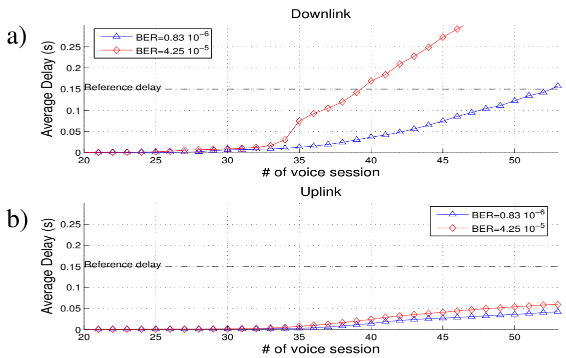

Fig. 3. VA scheme. Average voice packet delay vs. number of voice sessions; a) Downlink b) Uplink.

a maximum reference delay budget, i.e. 150 ms, which is also the maximum recommended tolerable one way end-to-end delay in order to maintain conversation between 2 parties [15]. Clearly, the speech quality depends also on the path over the wired trunk of the network, which here is not considered. We assume that, compared to the downlink, the wired path delay or the uplink delay has a smaller impact on the overall delay compared to the downlink delay. For larger delays on the wired path the voice capacity figures should be scaled accordingly. Also, we assumed that the *ACK* and broadcast frames are transmitted at 54 Mb/s data rate under a full ERP-OFDM mode.

Packet error events are considered by assuming a simple binary symmetric channel, where bit error rates occur independently according to a given BER. Such a model for the channel errors, though oversimplified, let us explore the sensitivity of VA and M-M schemes to a noisy channel. Also, we neglected the effects due to the path loss [12] since the distance from the AP is assumed small.

For the sake of completeness, we resume as follows the performance of the plain DCF scheme: up to 28 voice sessions, with low BER, the system is stable and downlink packets arrive well within the allowed delay budget. Also, the uplink delay is insensitive both to the number of voice stations and to the BER. Adding one more session, i.e. the 29-th, the system becomes unstable. This is the typical behavior as reported in literature [5].

We measured the gain obtained by the VA and the M-M cross-layer schemes. M-M has a large gain compared to the plain scheme but shows an abrupt transition to instability. Such a threshold effect, for the M-M, is due to the increase of the uplink delay, because the uplink enters starvation.

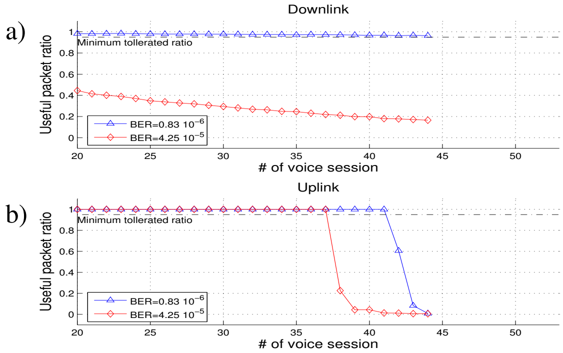

Fig. 4. M-M scheme. Ratio of useful received packets.

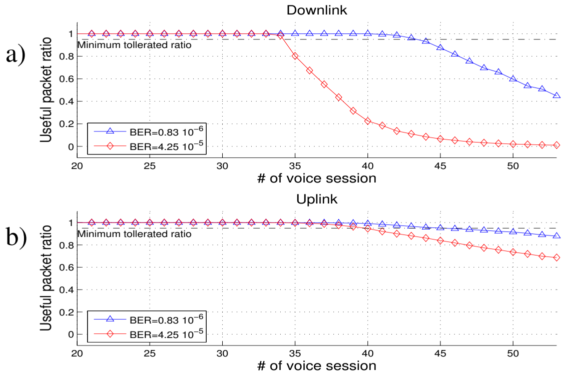

Fig. 5. VA scheme. Ratio of useful received packets.

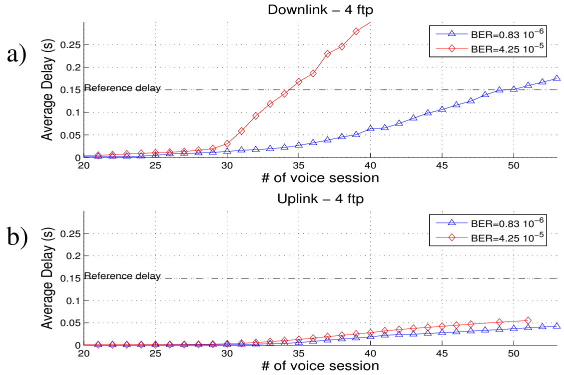

Fig. 6. VA scheme. Average voice packet delay vs. number of voice sessions; a) Uplink b) Downlink.

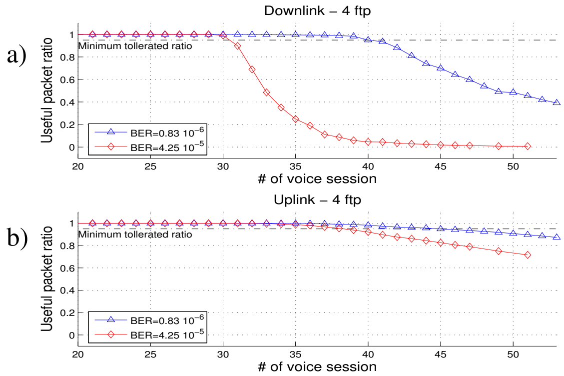

Fig. 7. VA scheme. Ratio of useful received packets.

The downlink still is served on average every T seconds. We noticed that (due to the multiplexing gain at the AP) M-M provides small delays and the system is stable up to around 41 voice stations.

Compared to the previos solutions, the VA scheme experiences a much more graceful degradation, as depicted in Fig 3 and Fig 6 and the limit of stability of the system is larger than the two previous schemes.

The improvement in the voice capacity is shown in the graphs reporting the ratio of useful packets, Fig. 4 and Fig 5. The M-M scheme sustains 41 stations under low BER, but is fragile for higher BER since the broadcast downlink is not reliable. The VA scheme sustains 43 sessions for low BER values, whereas higher BER values cause a capacity drop to 34 voice sessions; overall, the relative improvement of VA compared to the plain scheme is 53% and 30%, respectively.

We then measured the degradation introduced by alien data traffic. In the case of the VA scheme, Fig 7, 4 concurrent FTP sessions cause a drop of roughly 3/4 stations in the voice capacity. Such a smooth degradation in presence of alien traffic is due to the fair share of the medium provided by the IEEE802.11 MAC.

The M-M scheme was tested in the same conditions. The loss in performance is quite significant: as depicted in Fig 9 4 FTP sessions cause a drop from 41 to around 32/33 voice sessions. As captured in Fig 8, close to limit of stability, there exists a strong coupling of the M-M voice with FTP flows causing a larger downlink delay. Such a coupling effect is smaller for higher BER values: we ascribe this behavior to the fact that the congestion control of TCP reacts earlier due to channel losses and the delay ramp at the downlink is smooth, whereas, in the case of lower BER, downlink voice packets experience higher delays. For large number of voice sessions, FTP sessions decrease the transmission rate, and the downlink is guaranteed again a packet transmitted on average every T seconds.

Since non TCP-friendly flows pose some threats to the background data traffic we made a cautionary check that M-M and VA, do not improve voice capacity at the detriment of data traffic. FTP sessions share the AP incoming link (5Mb/s) with the active voice sessions: as in Fig 10, they occupy the residual bandwidth. The drop of aggregated throughput experienced by the FTP session concurrent with the M-M scheme, Fig 10b, corresponds to the sudden transition from stability to instability of uplink voice sessions. In the case of VA the degradation is more graceful since the system is stable over a larger interval.

VI. CONCLUSIONS

In this paper we showed that cross-layer techniques enhance significantly the voice capacity of IEEE802.11 infrastructured DCF networks. We compared the VA and the M-M schemes. Such schemes need minor changes in the customary configuration. The VA and the M-M schemes feed the MAC layer through compounding modules. VA, the scheme proposed in this paper, aggregates voice packets of the same VoIP flow and deliver them in the same frame. VA proved to relieve congestion at the AP using a plain IEEE802.11 MAC. Also the M-M scheme proved able to relieve the bottleneck of infrastructured DCF regulated WLANs at the AP. VA, in particular, shows a more graceful degradation at the increase of the disturbances on the wireless link, and milder transition from stability to instability compared both to the plain and the M-M scheme. Furthermore, a cautionary

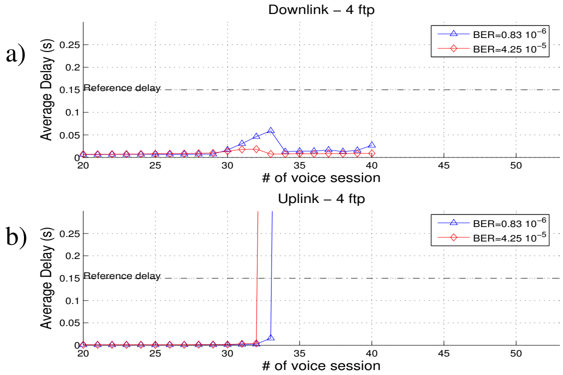

Fig. 8. MM scheme. Average voice packet delay vs. number of voice sessions; (a) Uplink (b) Downlink.

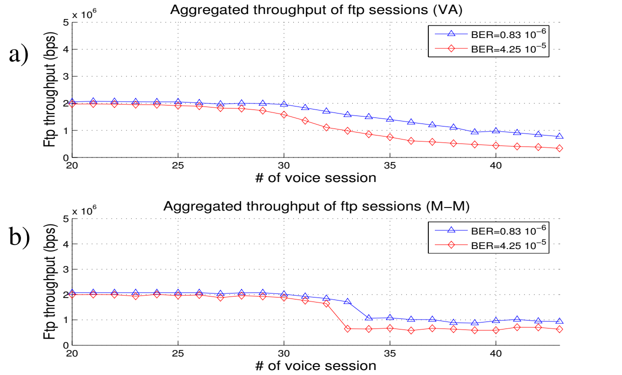

Fig. 10. Effect of the number of voice sessions on aggregated FTP throughput.

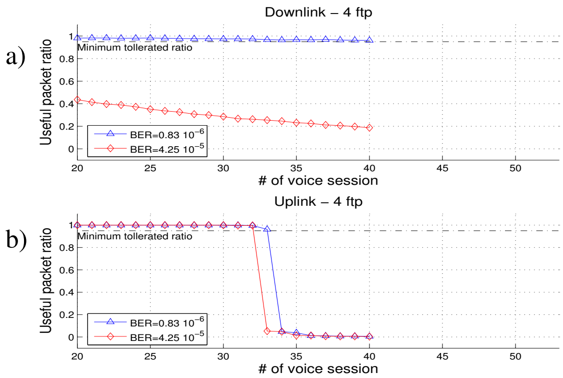

Fig. 9. MM scheme. Ratio of useful received packets.

check ensured that FTP regulated traffic is not starved by the VA and M-M schemes. IEEE802.11 schemes proved more robust to the presence of background traffic than the M-M scheme. Conversely, a point in favor of the M-M scheme the small delay per packet provided. Nevertheless, at higher transmission modes, we found that a large fraction of the gain in voice capacity is lost due to the lack of a reliable broadcast mechanism in IEEE802.11. In order to make the M-M scheme more robust, one direction is to trade off voice capacity for appropriate protection.

We envision two opposite research directions on the VA scheme. On one hand, we could employ RTP and cross-layer signaling: with this solution, the number of aggregated packets should be based on the estimate of the service time at the MAC layer, but the technique would not require major changes at the link layer. On the other hand, some hardware producers claim that the vertical aggregation is performed on their IEEE802.11 boards [16]: this would pave the way for a full VA based VoIP over IEEE802.11 architecture, with the performance gain outlined in this paper.

REFERENCES

[1] F. De Pellegrini, F. Maguolo, A. Zanella, and M. Zorzi, "A Cross-Layer Solution for VoIP over IEEE802.11," in *proc. of WPMC*, Aalborg, Denmark, 18-22 Sept. 2005.

[2] W. Wang, S. C. Liew, and V. O. K. Li, "Solutions to Performance Problems in VoIP Over a 802.11 Wireless LAN," *IEEE Trans. on Vehicular Tech.*, vol. 54, no. 1, pp. 366–384, January 2005.

[3] IEEE LAN MAN Standards, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications High-speed Physical Layer in the 5 GHz Band"*, ANSI/IEEE Std., September 1999.

[4] S. A. Baset and H. Schulzrinne, "An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol," Columbia University, Tech. Rep., 2004.

[5] S. Garg and M. Kappes, "Can I add a VoIP Call?" in *proc. of IEEE ICC*, Anchorage, 11-15 May 2003.

[6] ——, "On the Throughput of 802.11b Networks for VoIP," Avaya Lab. Research, Tech. Rep., 2002.

[7] K. Medepalli, P. Gopalakrishnan, D. Famolari, and T. Kodama, "Voice Capacity of IEEE 802.11b, 802.11a and 802.11g WLAN Systems," in *proc. of IEEE GLOBECOM*, Dallas, Texas, 29 Nov.-3 Dec 2004.

[8] M. M. Carvalho and J. Garcia-Luna-Aceves, "Delay Analysis of IEEE802.11 in Single-Hop Networks," in proc. of ICNP, Atlanta, Georgia, 4-7 November 2003.

[9] *Support of Voice Services in IEEE 802.11 wireless LANs*, Anchorage, 22-26, April 2001.

[10] A. Kopsel and A. Wolisz, "Voice Transmission in an IEEE 802.11 WLAN Based Access Network," in proc.of ACM WoWMoM, Rome, July 2001.

[11] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP," in proc. of IEEE ICC, Paris, 20-24 June 2004.

[12] M. Coupechoux, V. Kumar, and L. Brignol, "Voice Over IEEE 802.11b Capacity," in *ITC, Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, Antwerp, Belgium, 31 Aug. - 2 Sept. 2004.

[13] J. Yu, S. Choi, and J. Lee, "Enhancement of VoIP over IEEE 802.11 WLAN via Dual Queue Strategy," in proc. of IEEE ICC, Paris, 20-24 June 2004.

[14] The Network Simulator - ns-2. Available online at `http://www.isi.edu/nsnam/ns/`.

[15] *Recommendation G114, Enhancement Transmission Systems and Media – General Characteristics of International Telephone Connections and International Telephone Circuits*, ITU-T Std., Feb. 1996.

[16] US Robotics - `http://www.usr.com/download/whitepapers/125mbps-wp.pdf`.