Research Article

A Comparative Survey of Adaptive Codec Solutions for VoIP over Multirate WLANs: A Capacity versus Quality Performance Trade-Off

A. Sfairopoulou, B. Bellalta, C. Macian, and M. Oliver

NeTS Research Group, Universitat Pompeu Fabra, 08018 Barcelona, Spain

Correspondence should be addressed to A. Sfairopoulou, anna.sfairopoulou@upf.edu

In multi-rate WLANs, users can suffer transmission rate changes due to the link adaptation mechanism. This results in a variable capacity channel, which is very hostile for VoIP and can cause serious quality of service (QoS) degradation in all active calls. Various codec adaptation mechanisms have been proposed as a solution to this, as well as to solve congestion problems on WLAN environments. Here, these solutions are presented, categorized according to the adaptation policy and scenario they implement, and evaluated at call-level in terms of the resulting blocking and dropping probabilities, as well as the perceived voice quality. To define a common performance metric, a new index named VGoS-factor is presented, which, by combining these capacity and quality indicators, can provide an overall view of the capacity versus quality trade-off of the proposed mechanisms and consequently help in choosing the adequate policy for each scenario.

1. Introduction

VoIP over WLANs has become a hot research topic during the past years due to the widespread deployment and ease-of-use of both technologies. Nevertheless, specially the commercial deployment of VoIP over WLANs cannot be achieved unless most of the VoIP quality of service issues are solved efficiently. The 802.11 standard [1], as most of the IP networks and their underlying technologies, was not created having voice in mind and therefore severely limits the successful deployment of VoIP. Voice transmission over the wireless link under variable channel conditions can easily suffer from an increased packet error and loss ratio, with direct effect on its performance and quality. In spite of all the research efforts in this area, there are still unsolved issues concerning VoIP calls, mainly caused by the specific wireless network characteristics. One of the most problematic among them is the multirate transmission.

Multirate transmission is one of the key features of the IEEE 802.11 PHY/MAC layer which allows each mobile node to select its physical layer parameters (modulation and channel coding) to optimize the bit transmission over the noise/fading-prone channel. These unexpected rate changes occur on the mobile nodes due to a reaction of the link adaptation (LA) algorithm of the 802.11 specification to a number of different circumstances, such as user movement, obstacle interference, and so forth. Heusse et al. [2] identified an anomaly caused by the LA mechanisms, mainly originating at the 802.11 "fair" channel access mechanism. They observed that the rate change of *one* of the nodes of the cell causes a general degradation on the transmission rate of *all* the other nodes. In other words, these unexpected rate changes have an impact on the transmissions of all active nodes and produce a general degradation of the network performance and a very hostile environment for VoIP. The reason behind this is that by reducing its transmission rate, a node demands more cell resources (channel occupancy time) in order to transmit the same amount of data. This increased demand can lead the system to a congested state (caused by less resources available to the system than the ones required) with a direct impact on all active calls.

Various studies [3–5] have demonstrated that the actual capacity of an 802.11 cell in terms of simultaneously active calls is surprisingly low taking into consideration the high data rates that the different standards of this technology can offer. Due to this limited voice capacity, a careful planning of the number of calls accepted, and under which conditions these will be accepted, is mandatory in order to maximize the cell capacity and distribute efficiently the network resources among them. This is where call admission control (CAC) methods enter, trying to determine the available cell capacity at each instant and accept or reject new calls accordingly [6]. However, this task's complexity increases further when that capacity becomes variable due the multirate transmission effect. The CAC algorithms alone are not able to deal with this situation, since it affects calls *already admitted* to the network.

Additionally, although a number of enhancements to the IEEE 802.11 were proposed, like the well-known EDCA mechanism for traffic differentiation, these alone cannot adequately address the multirate performance anomaly when it affects traffic flows of the same traffic class, which is the case studied here. In this pure VoIP scenario, all packets arriving to the MAC layer have the same priority and use the same access category queue. Even if it was possible to schedule each traffic flow independently, similar to the mechanism proposed by Garroppo et al. in [7], this would result in the starvation of some calls (e.g., those which are causing the congestion) until they are dropped by the users, as the amount of offered traffic is higher than the resources available. Nevertheless, EDCA is a required piece in the overall system in presence of heterogeneous traffic scenarios, where for example the VoIP calls suffer starvation due to the presence of elastic traffic such as from TCP flows. In this case, a properly designed tuning algorithm using the EDCA parameters have to be used to isolate the VoIP calls from the TCP flows. It would still be required, however, a different mechanism to solve the multirate performance anomaly of the VoIP calls.

Having this in mind, a different solution should be used to avoid unnecessary call dropping, such as reducing the traffic load of the WLAN to acceptable values. In a VoIP scenario, this could be done by changing the voice codecs of one or several calls to new ones with lower overall bandwidth requirements. Various such codec adaptation mechanisms have been proposed in the literature [8–10], including the authors' own work [11, 12]. These proposals are reviewed briefly in the next section. Their concrete implementation issues are not extensively discussed here since the comparison focuses on the different "strategies" they use, which we categorize under the name of *decision policies*, presented in Section 3. Thus, the main purpose of this paper is to study how the various codec adaptation solutions can perform under different usage scenarios and following one of the possible decision policies for the adaptation procedure.

At the same time, various call admission control mechanisms are commonly used in WLANs to control the access to the limited resources. These mechanisms only accept/reject new incoming calls based on the available bandwidth at the moment the request is done [6, 13]. They are unable to react to the sudden transmission rate changes occurring while an accepted call is active and to the variations on the total capacity that this rate change implies.

Figure 1 shows a schematic approach of a generic codec adaptation solution and the parts analyzed in this study are highlighted. How the codec adaptation mechanism is implemented and how the codec negotiation and management takes place is outside the scope of this paper, for it is protocol and technology specific. One of the available solutions is the use of the SIP [14] protocol, which provides the possibility of updating the codec used during an active connection by means of an additional INVITE message and while maintaining the voice session active. This would require the presence of a SIP proxy and some software-based modification at the access point. Nevertheless, the focus of this paper will fall only on analyzing and comparing, in terms of performance, how the different solutions address the capacity versus quality trade-off according to the adaptation policy they use, as presented in Section 3, and to a comparison among these policies under different scenarios. The results provided in this paper can thus serve as a basic reference for future research on this area.

In order to facilitate the comparison between the solutions, a new voice GoS (grade of service) metric, the VGoS-factor, is proposed (Section 4), which can unify quality and capacity metrics for VoIP flows in a WLAN and can guide the decision on when, how, and where to apply codec adaptation solutions. A performance analysis of the different codec adaptation solutions is provided in Section 6, focusing on the trade-offs given by the different policies in the overall performance of the proposal. The obtained results and conclusions are valid for the generic scenario considered in this paper, although the conclusions are general and are easily adaptable to other scenarios and/or technologies which suffer from the multirate anomaly. Finally, the main conclusions and open issues are discussed in Section 7.

2. Codec Adaptation Solutions

Adapting the parameters of VoIP flows (like codec or packetization interval used) has been considered before and similar works can be found in the literature. An example is the work presented in [8], where McGovern et al. evaluate whether a rate change generates congestion or not and then propose a codec change for the node that suffered the rate change. To choose the new codec, they use a metric called channel occupancy time, which represents a fraction of the channel time per unit time required by a full-duplex VoIP call for a given codec and rate. Only the node that suffered the rate change is adapted, which in some cases is not sufficient as we will see later in our evaluation. The work presented by Chen et al. [9] uses the same basic idea (adjusting codec and packetization interval) in order to solve capacity problems mainly occurring due to the handoff procedure. By lowering the quality of some of the calls in the network, it can adjust the distribution of resources among existing calls and permit others to enter. In fact this is an example of how a codec adaptation algorithm can work together with a call admission control technique in order to provide a more
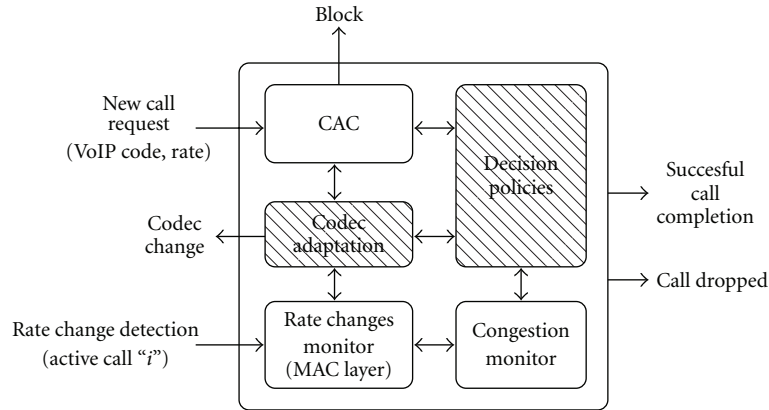
FIGURE 1: General codec adaptation solution.

efficient use of the network resources. In [15], Garg and Kappes propose to reserve some of the network capacity so as to be able to absorb a data rate change for some of the connections. Clearly this implementation can be very wasteful of network resources, especially if the reservation is such so as to deal with rate drops from 11 Mbps to 1 Mbps. A simpler solution would be to drop the problematic call as proposed by McGovern et al. in a different paper [10]. Although this way the problem is solved very fast, dropping the call has a negative effect on the user's opinion of the network's QoS.

In Trad et al. [16], when the channel is detected congested (based on Real-Time Control Protocol (RTCP) packet loss and delay feedback information) a central element performs the adaptation of codecs using common transcoding methods for all calls *entering* the wireless cell (incoming flows only). The main assumption is that the AP is acting as a bottleneck and thus, a solution is needed on the border gateway at the entrance of the network. Their proposal is dealing with network congestion rather than the multirate anomaly. Kawata and Yamada [17] also propose double transcoding performed at the AP, both for incoming and outgoing calls, in order to adapt to the new capacity varied by the multirate mechanism and thus achieve the same effect as a codec change, but without actually renegotiating a codec with the other end of the call. Both of these proposals are based in transcoding which can be very expensive in terms of processing power needs. Additionally, there are not any real codec changes since the AP transcodes the VoIP calls. For this reason, they will not be further considered in the performance comparison of the different solutions.

Other proposals based on varying the scheduling algorithms in the PHY/MAC protocol itself can be found, like [18]. We are not further considering these solutions since they can be more difficult to implement and to provide the necessary backward compatibility with the already deployed equipment, as they require modifications at the standard operation of 802.11.

In [19], Tebbani and Haddadou try to guarantee the QoS of high priority calls against the low priority ones (where priority is determined by the SLAs agreement that users establish with providers) by "down-switching" codec and packetization interval of one or more low priority calls based on a set of information (call and user ID, call codec, session duration, packet loss and end-to-end delay, etc.). Their focus is not on the multirate effects but on the congestion provoked by the existence of additional VoIP sessions on the cell, and the algorithm focuses on adapting existing low priority calls in favor of the high priority ones.

To alleviate the QoS degradation of VoIP calls due to the multirate anomaly, a voice codec adaptation (VCA) solution has been presented by Sfairopoulou et al. [11, 12], based on cross-layer feedback from RTCP packets and the PHY/MAC layer. The feedback consists of the voice QoS information that can be obtained from the periodic RTCP packets, as well as the announcements of rate changes arriving from the MAC layer. These parameters are constantly monitored in real-time for each call and can trigger a recovery mechanism to adapt the voice codec that is being used, to one that fits better the new channel conditions and can lower the congestion levels of the cell. The codec change can affect just the call that has changed its transmission rate (*slow* call) or some of the other active calls as well, in case the first change is not enough. All the codec changes are performed without interrupting the call and thus avoiding any call dropping. Note that the VCA solution can be implemented entirely at the application layer, without requiring any modifications at the MAC and PHY layers of the IEEE 802.11 standard.

This work has been used as a basis for the work presented in [20], by Tüysüz and Mantar. The authors focus on a combination of codec change and packetization interval change (frame size) based again on MAC information on rate changes and RTCP feedback for quality problems. The main difference lies in classifying packet losses on the ones due to congestion or due to error prone channel; in case of congestion due to the multirate anomaly, using codec change on the call that changed rate and incrementing the frame size of others can be helpful, while in case of error-prone channel, using bigger frame size will cause more packet losses so it should be avoided.

By combining the VCA algorithm with a CAC in [21], Sfairopoulou et al. in an extension of their previous work address the limited capacity issue while at the same time controlling the multirate effects. This combined CAC/VCA solution would allow, for example, an incoming call to be admitted in an otherwise saturated network, by accepting to use an "unfairly" low-rate codec for its transmission rate or by reducing the VoIP codec of *other* active calls, in order to optimize overall resource usage and capacity. A set of guidelines on how to tune the CAC/VCA solution are also presented in [21].

A categorization of the most relevant of the presented solutions according to the decision policy and usage scenario is provided next, in Section 3. We will not consider further the solutions that use transcoding and resource reservation, as the focus is on codec adaptation procedures.

3. Policy-Based Codec Selection and Usage Scenarios in WLAN Hotspots

Every codec adaptation algorithm can be considered to follow a *decision policy*, which dictates the usage of voice codecs and the manner that the adaptation is performed, or in other words, if and how many calls should change codec, depending on the criterion to be maximized. For example, if the main objective is to obtain the maximum average MOS (mean opinion score) value (E[MOS]) a valid policy would be "always use the G.711 VoIP codec, independently of the considered rate, for it gives the best voice quality". There is a clear trade-off between the different criteria, especially between the achieved quality and the quantity of accepted calls. The analysis of this trade-off becomes a difficult task when a larger number of metrics are taken into account, which has motivated the use of a new combined metric, the VGoS, that will be introduced and described in the next section.

One of the additional basic parameters to consider is the usage scenario of these policies. In this study, we consider three different usage scenarios where adaptation can be applied: only in the case of *rate changes*, only at the entrance of *new calls*, or on both previous cases (*rate changes and new calls*). The first usage scenario is oriented on solving the multirate effect and decreasing the dropping ratio, maintaining an acceptable QoS of the ongoing calls. The second scenario is focused on maximizing the capacity of the cell by decreasing the blocking ratio and accepting more calls, by lowering their codec. Finally, the third scenario is a combination of the other two.

The term policy refers basically to whether codec adaptation is applied on none, one or a number of calls (i.e., nonadaptive when no codec adaptation is done, single-adaptive when only one call is adapted and multiadaptive when more than one call change their codec). A more extensive work on decision policies can be found in [22]. In this paper we use the most representative among them with the main purpose to describe and evaluate the performance of the proposed solution. These policies are thus divided in three subcategories.

(i) *Nonadaptive Policies.* Under a fixed-codec policy, each call maintains the initial codec throughout its duration. No codec adaptation is performed and when any congestion occurs a simple drop/block solution is applied. Using this policy, calls can start with any codec from the set of supported codecs $\mathcal{C}$. In the general case, this policy produces the maximum number of call drops in a multirate environment, since no adaptation is performed. The work of McGovern et al. [10] can be considered a concrete implementation of this policy (i.e., no codec adaptation is considered, thus directly drop the call that changes rate when this change causes congestion) and will be used as such for comparison in Section 6.

(ii) *Single-Adaptive Policies.* In this category, the only calls that change their codec are the ones that actually created the capacity variation state. This includes both new calls trying to enter and finding the cell capacity already fully used, as well as nodes that change to a lower data rate in the middle of a call causing a decrease on the total cell capacity. These policies can be considered the fairest ones, since the only call affected by a codec change is the one that created the problem. The idea is to admit/not drop a call if it agrees to enter/continue using a codec lower than the one requested (c'), if it is supported by the mobile node. This policy allows each VoIP call to proceed with a worse voice quality but reducing its blocking and dropping probability, as it is able to check if lower rate codecs would be better suited to the network conditions and new capacity. Considering the special case where this policy is applied only as a remedy for rate changes, many similarities with the work of McGovern et al. in [8] can be found, where the authors propose a solution based on codec adaptation of only the affected call in the case of a rate change.

(iii) *Multiadaptive Policies.* Using these policies, in order to allocate space for new calls or to solve the channel congestion caused by rate changes, any number of calls can be requested to change their current codec to a lower one (c'), if it is supported by the STA. The new call (or the one that suffered the rate change) will be the first to enter the adaptation procedure as in the single-adaptive policy in order to achieve some fairness; however other calls can be also requested to change their codec as well, until the congestion is solved. It is a more complex solution compared with the non- and single-adaptive policies but also the one that gives best results in terms of dropping and blocking rate. There are different criteria that can be used to select which calls have to suffer a VoIP codec degradation, like the priority of the call, its duration, the codec used, and so forth. All of this criteria can create different subcases of the multiadaptive policies. For the sake of simplicity only one possibility is presented here: choose the next call

for adaptation randomly from all the active calls on the cell. If this call cannot switch to a lower codec (or if its change is not enough to free the necessary capacity), the algorithm will choose a new one (or an additional one, since more than one call can change codecs). This process is repeated until the congestion is solved or until no successful combination of codecs can be found after all active calls have been checked. Since the CAC/VCA module keeps complete information about the cell and the active calls in it, it iteratively performs the calculations stated above, as well as computing the resulting cell state, prior to *actually* triggering the codec changes. Hence, if no feasible solution is found, the problematic call will be dropped or blocked, without affecting the others. The work of Chen et al. [9] can be considered a concrete implementation of this policy acting only at the entrance of new calls (scenario "on new") since any number of calls can change their codec for new ones (handoff calls in their study) to enter. Furthermore, the VCA algorithm of Sfairopoulou et al. [11] is also an implementation of this multiadaptive policy but for the case of rate changes (scenario "on rate") since it focuses on the effects of the multirate anomaly. The same applies for the proposal of Tüysüz and Mantar [20], with the addition of packet size adaptation as well. Finally, the combination of codec adaptation with a CAC algorithm (Sfairopoulou et al. [21]) can be considered as an implementation of the multiadaptive policy for the "on both" scenario, since it intends to minimize the call dropping due to the multirate effect while maximizing the admitted calls at the cell. The performance of all these solutions will be extensively analyzed in Section 6.

In Table 1, a comparison of the most relevant works discussed here is provided, categorized according to the *policy* they use and their usage scenario.

4. The VGoS-Factor, a Capacity versus Quality Trade-Off Metric

From the policy analysis we can realize that making a choice over which decision policy is the best in absolute terms is a difficult task, since achieving a high value in all of the system performance metrics under study is quite complicated. Although the usual trade-off between quality and quantity is to be expected here as well, in order to provide some kind of qualification for the set of policies and decide where and how to use codec adaptation depending on the traffic load, a new grade of service (GoS) factor which combines the above metrics is needed. This "quality and quantity" factor, called VGoS-factor, can provide a combination of blocking (B) and dropping rate (D) with the average normalized MOS (MOS_n) in the hotspot: the first two serving as quantitative metrics of the system's capacity to accept new calls and finalize correctly the accepted ones, while the third one as a speech quality metric.

TABLE 1: Comparison of codec-based solutions to multirate anomaly.

Usage scenario	Decision policy		
	Non adaptive	Single adaptive	Multi adaptive
On new calls	—	—	Chen et al. [9]
On rate changes/on congestion	McGovern et al. [10]	McGovern et al. [8]	Tüysüz and Mantar [20]
			Sfairopoulou et al. [11]
			Tebbani and Haddadou [19]
On new calls + rate changes	Garg and Kappes [15]	—	Sfairopoulou et al. [21]

Call admission control in cellular networks is designed trying to guarantee both grade of service at call level and quality of service at packet level [23]. Unfortunately there is not one single metric that can combine them both and provide an overall view of the performance of a CAC mechanism and consequently of the cell capacity itself. The VGoS-factor is a system performance metric that tries to respond to this demand by combining both dropping and blocking feedback as GoS parameters at call level with the packet loss and delay/jitter metrics (reflected in general in MOS or equivalent voice QoS metrics like the E-Model [24]) representing the QoS at packet level.

Grade of service is the probability of a call to be blocked, either entering the cell as a new request or as a handoff call from another cell. In [25], GoS is calculated as

$$GoS = \alpha \cdot P_h + P_n, \tag{1}$$

where P_h is the handoff failure probability and P_n the new call blocking probability, while $\alpha = 10$ is a weight parameter to represent the penalization for dropping a handoff call relative to blocking a new call.

In the scenario used here, no handoff calls are considered (the VGoS can be easily extended to consider handoff calls; however, a handoff call not accepted in the cell has a similar effect to a dropped call. In this paper we only want to focus on the call dropping due to rate changes) and the dropping ratio refers only to the already active calls in the cell, so the above GoS can be modified by simply replacing P_h with this dropping ratio. Considering this change as well as the incorporation of the MOS value, the VGoS-factor is calculated as

$$VGoS = M \cdot (\alpha \cdot D + B), \tag{2}$$

where B and D are the blocking and dropping probabilities and M is calculated as the distance of the estimated average MOS from the maximum MOS ($MOS_{max} - E[MOS]$), with $MOS_{max} = 5$. The VGoS metric extends the traditional GoS by multiplying it with the distance of the estimated average MOS from the maximum MOS. The decision to include M as a product is due to the following reasons: (1) it is a completely different metric and does not refer to ending

Table 2: Codecs parameters.

$\mathcal{C} = [$G.711, G.726, G.729, G.723.1$]$			
Codec	Bitrate (B)	Packet size (L)	MOS
G.711	64 Kbps	160 B	4.1
G.726	32 Kbps	80 B	3.85
G.729A	8 Kbps	20 B	3.7
G.723.1	5.3 Kbps	20 B	3.6

a call, as the blocking and dropping probabilities do, and (2) the achieved average MOS is seen as a proportionality (scale) factor to traditional GoS. This is, given two policies with the same blocking and dropping probabilities, the policy with a better $E[MOS]$ will show the best VGoS (that is the lower VGoS, since the goal is to minimize it, similar to the traditional GoS). Additionally, introducing the M as a product, we are not changing the relation between the blocking and dropping probabilities, and the M becomes simply a scale factor.

Notice that $E[MOS]$ can be computed easily in real time (e.g., using the E-model [24]) from the feedback of the voice packets transmission. The computation requirements of E-model are low [26] and thus it could be easily embedded in the WLAN access point as a part of the considered solution. For the simulation experiments we set $\alpha = 10$ as in the initial GoS. As a general rule, we have considered that users have a stronger reaction against call dropping than against lower call quality, and we have adjusted the VGoS factor accordingly.

5. WLAN Scenario for VoIP Calls

A scenario which consists of an IEEE 802.11e hotspot that includes a QoS-enabled AP (QAP) and a number of QoS-enabled mobile stations (QSTA) is considered, as depicted in Figure 2. The standard IEEE 802.11 [1] and 802.11e [27] parameters are used at MAC and PHY layers, with the available set of transmission rates being $\mathcal{R} = [11, 5.5, 2, 1]$ Mbps (considering the case of the 802.11b PHY specification, although the conclusions presented here are valid—properly scaled—for other PHY specifications such as the 802.11g or even a mixture of "b" and "g" STAs). Only bidirectional VoIP traffic is considered, where each active call can use different voice codecs (and thus different bandwidth requirements) and different transmission rates. An important point is that not all VoIP STAs are always able to participate on the codec adaptation process as their VoIP phone may not support multiple VoIP codecs and/or the end-users may not be willing to use this procedure. We reflect this issue considering also how the percentage of participating nodes affect the codec adaptation process. EDCA is used at the MAC layer, although its traffic differentiation and prioritization capabilities are not considered since there is only voice traffic in the cell. The inclusion of elastic or other types of rigid flows will be considered in further work.

The basic parameters considered for the set of voice codecs used are shown in Table 2. The standard packet size

achieved with 20 ms (30 ms in G.723.1) packetization interval is used for all policies. Only when studying the proposal of Tüysüz and Mantar [20] which adapts both codecs and packet sizes of flows we use a double packetization interval (and thus double the packet size).

5.1. A VoIP Upper-Bound Capacity Indicator for Multirate/Multicodec WLANs.
In [28], Hole and Tobagi proposed a capacity index as a reference upper bound for the number of maximum accepted calls N in a 802.11 cell (using the DCF). They assume that all nodes transmit at the same rate and use the same voice codec.

This capacity indicator is based on some arguably fragile assumptions, such as no collisions and no link errors. However, as shown in [6], the capacity of a WLAN cell is basically limited by the saturation of the AP (it is not able to deliver as many packets as required), resulting in the starvation of the downlink VoIP flows. Surprisingly, the AP becomes rapidly saturated for an unexpected low number of active VoIP calls, mainly due to the high traffic load that this has to deliver (all downlink VoIP flows pass through the AP). Thus, when the AP saturates, there are in fact few competing STAs and the collision probability remains low. On the other hand, as it is also shown in [28], the channel errors have a remarkable impact on the reduction of the cell's capacity. However, since we consider a multirate scenario and thus that a link adaptation mechanism is used, the impact of the channels errors can be minimized: the system will choose a lower PHY rate if the channel is not able to satisfy a minimum packet error ratio at the current rate.

Since our study focuses on a multirate WLAN and nodes can use any codec chosen from a set of available codecs, we have modified this first index to produce a new multirate/multicodec capacity index. We will describe these two indexes here in detail.

(a) Upper bound index of call capacity of an 802.11 network, by Hole and Tobagi.

In their work in [28], Hole and Tobagi have provided a mathematical upper limit of an 802.11 network capacity of VoIP calls (in a scenario with only VoIP calls). In this scenario the only constraint for admitting (or not) a new call is the available throughput. Assuming that all calls use the same codec and the same link rate, the upper bound of the value of maximum accepted calls N is given by the equation:

$$N = \left\lfloor \frac{1}{B/L[2 \cdot T_s(R^d) + (T_{\text{slot}} \cdot (\text{CW}_{\min}/2))]} \right\rfloor, \quad (3)$$

where T_s is the time required to transmit a VoIP packet, calculated as

$$T_s\left(R^d\right) = \left(H + \frac{H_{\text{mac}} + L_{\text{data}}}{R^d}\right) + \text{SIFS} + \left(H + \frac{L_{\text{ack}}}{R^b}\right) + \text{DIFS}, \quad (4)$$

where H is the time to transmit all the PHY headers (preambles) at $R_{\text{PHY}} = 1$ Mbps, R^d is the data rate used from the set of data rates $\mathcal{R}^d$, and R^b is the basic rate used from the set of basic rates, $\mathcal{R}^b$. Notice the dependence of T_s with the packet length (L_{data}) itself but also with the transmission
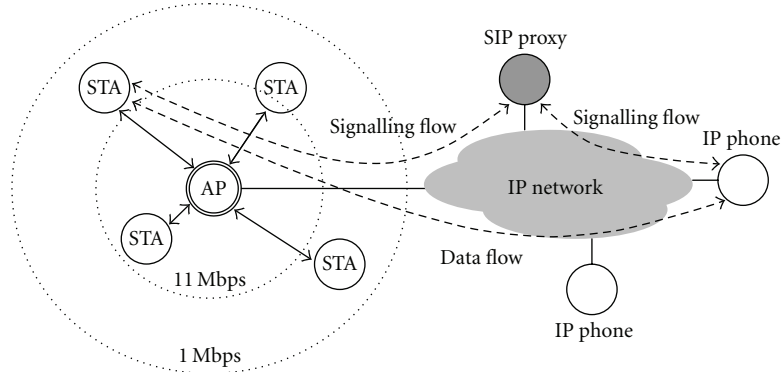
FIGURE 2: A VoIP WLAN Hotspot scenario.

TABLE 3: Component times of T_{VOICE} and T_{ACK} for 11 Mbps data rate.

T_{VOICE}	PLCP preamble & header	192.0 us
	MAC Header + FCS	20.4 us
	IP/UDP/RTP header	29.1 us
	Voice Data	(Voice octets $\times$ 8/11) us
T_{ACK}	PLCP preamble & header	192.0 us
	ACK frame	10.2 us

rate R^d used. In (3), B/L is the transmission packet rate for the VoIP codec used and T_{slot} is the duration of an empty SLOT. The default parameters used assuming a data rate of 11Mbps are given in Table 3. The default values for CW_{min}, SIFS, T_{SLOT}, and DIFS (refer to the 802.11 standard [1] for the definition of these parameters) are respectively 31 slots, $10\,\mu$s, $20\,\mu$s, and $50\,\mu$s.

(b) Modified multirate/multicodec capacity index.

Considering the existence of multiple transmission rates and VoIP codecs in our scenario, an extension of the capacity index proposed in [28] is presented. Using the same argument and assumptions as before, a state is feasible if it satisfies the condition

$$\sum_{R^d \epsilon \mathcal{R}} \sum_{c \epsilon \mathcal{C}} n\left(c, R^d\right) \frac{B(c)}{L(c)} \left(2T_s\left(c, R^d\right) + T_{\text{SLOT}} \frac{CW_{\text{min}}}{2}\right) \leq \rho,$$

(5)

where $n(c, R^d)$ is the number of calls using codec c and rate R^d, $T_s(c, R^d)$ is the time required to transmit a VoIP packet using codec c and rate R^d, $B(c)/L(c)$ is the transmission packet rate for VoIP codec c and ρ is the normalized time allocated for VoIP calls. Computing ρ with multiple traffic classes is not an easy task. This is because ρ depends on many nonlinear relations between the number and characteristics of the active traffic flows (e.g., refer to [29] to see how the 802.11 behaves with both elastic and rigid flows). It also depends on the traffic differentiation mechanisms considered (e.g., EDCA) and how these are configured [6]. All these facts are out of the scope of this paper where a pure VoIP scenario is considered, and therefore we consider

a constant ρ equal to 1, meaning that all transmission resources are assigned to VoIP traffic.

In order to validate the accuracy of this upgraded multirate/multicodec index (MR/MC), a comparison was performed for a number of different test cases against the results obtained from simulations using NS-2 [30]. In each case, given a fixed number of calls using a codec c_i and transmitting at rate R_i^d, the maximum number N of calls using codec c_j and rate R_j^d accepted in the cell without causing congestion has been calculated both by the MR/MC index and experimentally by the NS-2. The results give a 100% match between the two. We will use this modified MR/MC index as our voice capacity indicator for the simulation experiments at the flow-level described later.

5.2. Fairness Issues. This paper does not cover the fairness issues of the codec adaption solutions, meaning the case when calls which are not affected directly by the rate change have to lower their VoIP codec and hence, observe a somewhat worse quality in terms of MOS. The fundamental assumption, that this change would allow to maintain the system stability, is not completely valid as there is always the alternative to simply drop or block the call which is causing the instability. Thus, some incentives to the user which accepts to participate on the codec adaptation should be introduced.

6. A Comparative Performance Analysis

In this section, a comparative analysis between the solutions which apply codec adaptation and described in Section 2 is provided (recall Table 1). We consider the combination of decision policies with the three basic scenarios: applying codec adaptation only on rate changes, only on new call arrival, or on both rate changes and new call requests. We are interested in evaluating the call-level metrics, such as blocking and dropping probabilities, which basically require the simulation of the arrival and departure call processes, as well as the random capacity changes and the resultant cell capacity. The resources wasted for signaling and control packets (SIP and RTCP) are not considered. The paper focuses on a conceptual comparison of the different policies

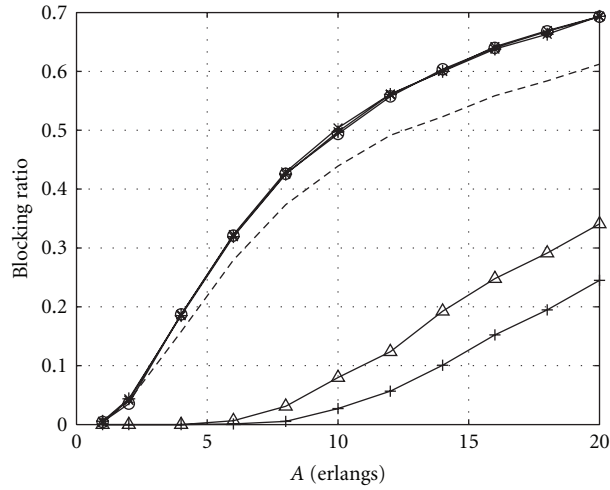which will allow their evaluation independently of these implementation details.

These codec adaptation schemes and policies have been evaluated through a C++ simulator, developed from scratch using the COST simulation toolkit [31]. Roughly, the simulator consists of three blocks: (i) a call generator, responsible for generating the new calls and choosing/assigning the characteristics of each call (STA's processor capacity, VoIP codecs supported, etc.), (ii) a rate change block, which randomly changes the rate of active calls, and (iii) the codec adaptation block which, based on the number of active calls and the current rate and codec of each one, decides the actions to be taken in the case of a rate change or a new call arrival. The WLAN cell capacity is forecasted using the multirate/multicodec (MR/MC) capacity index explained and validated in Section 5.1.

VoIP call requests arrive to the cell with rate λ following a Poisson process. The call duration follows an exponential distribution with mean equal to $1/\mu$ = 240 seconds, so that the offered traffic load A varies from 1 to 20 erlangs. A frequency γ of 1/100 rate changes per second and mobile node is considered unless other values are stated (as in Section 6.3). Using this value, a call which finishes successfully (it is neither blocked nor dropped) suffers in average two rate changes. The new rate is selected from the set of available rates following a uniform distribution, since it has been assumed that all rates have the same probability to be chosen. This is the worst possible case since, usually, changes are between adjacent rates, resulting in a more predictable situation. The consideration of a specific mobility model would produce more real results, leading however to only a particular case from which it would be difficult to extract general conclusions. Our main goal is to compare the different policies in rather ideal conditions, limiting at maximum the variables which impact on their performance and focusing only the sensitivity of the evaluated scheme on the traffic load, the rate change frequency, and the coexistence of participating and non-participating nodes in the codec adaptation process. The effect of the mobility models will be considered for further research.
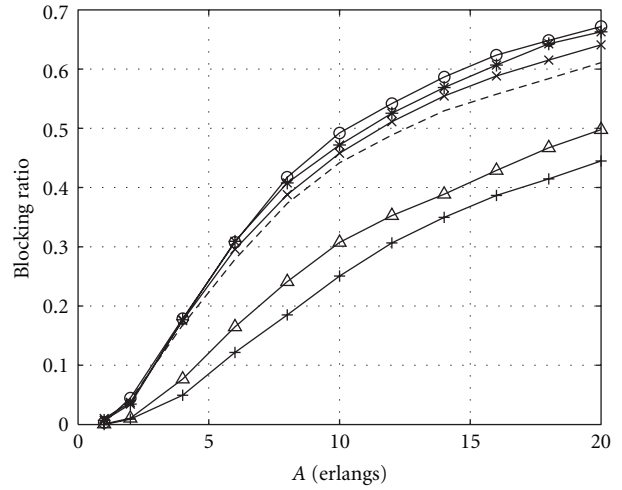
6.1. Comparison of Codec Adaptation Strategies. In Figure 3, the different solutions are compared. In the left column (Figures 3(a), 3(c), and 3(e)), we assume that all STAs participate in the adaptation procedure, they can support all codecs and start with the highest rate codec (*G*.711). The blocking (Figure 3(a)) and dropping (Figure 3(c)) probabilities (*BP* and *DP*, resp.) as well as the *E*[MOS] in the cell (derived from the theoretical values associated with each codec [32]) (Figure 3(e)) were obtained for each case and plotted against the offered traffic *A* erlangs. We also examine the case where only 40% of the STAs are participating and implement the adaptation solutions (right column Figures 3(b), 3(d), and 3(f)). The nomenclature used on the figures implies whether it is a nonadaptive (NA), SingleCodec adaptive (SC), or MultiCodec adaptive (MC and MCF—when frame size is also adapted) policy, in combination with the usage scenario

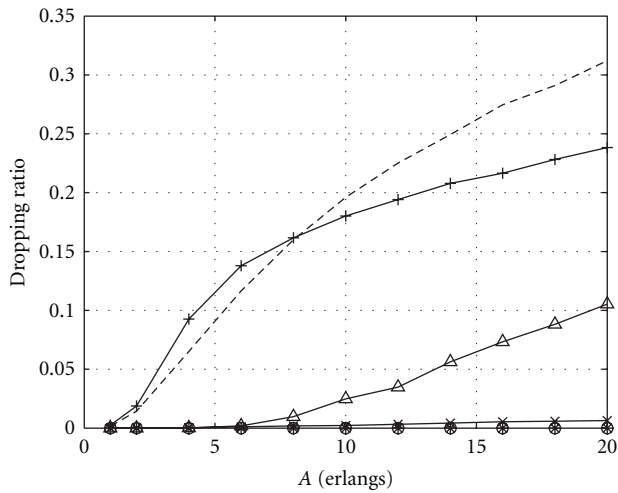(rate, new, rate+new). The main observations considering each scenario are as follows.

(1) *Nonadaptive Case (NA)*. This case presents high blocking and dropping probabilities leading to less active calls, as expected, since no adaptation is possible. At the same, time however, it gives the highest *E*[MOS] results since, although the number of successful calls is actually smaller, all of them use the highest possible codec through all their duration. Obviously, here the percentage of nodes participating on the adaptation has no influence.

(2) *Apply Codec Adaptation Only on Arrival of New Calls (MC-new)*. This case provides a high cell capacity in terms of number of calls, as it is able to adapt the VoIP codec of the active calls in order for more new calls to fit the current available capacity and avoid blocking them (it shows the lowest *BP*). On the other side, it presents a considerable high dropping probability, only equivalent to or lower than the nonadaptive case seen before. This is due to the fact that by using the solution of MC-new, more calls are admitted in the cell but with low bitrate codecs. Thus, they can support better the rate changes occurring afterwards (they have less capacity demands) than in the NA case where the calls, if admitted in the cell, use higher bitrate codecs. However, since no adaptation is performed on rate changes, the policy presents worst dropping results than the ones acting on the on rate scenario. At the same time, as traffic load *A* increases, the voice quality (*E*[MOS]) is progressively reduced since almost all new calls are asked to use the low bandwidth codecs in order to be accepted. A lower number of nodes participating in the adaptation process results in an increase of the blocking probability (there are less possibilities for codec adaptation) as seen in Figure 3(b), and also an increase of the dropping probability (Figure 3(d)) since there are now more nodes starting with the highest rate codec and the probability that a rate change will result in a call dropping increases.

(3) *Apply Codec Adaptation Only during Rate Changes and Use a Standard CAC for New Calls (Nonadaptive CAC, Block Any Call That Cannot Fit In) (SC-rate, MC-rate, MCF-rate)*. This case provides a high protection for the already accepted calls (very low dropping probability compared to the fixed policy) at the cost of blocking a high number of new call requests. Notice that all three solutions in this category (SC-rate, MC-rate and MCF-rate) have an almost identical performance since in the average case it will be enough for solving the congestion to adapt the codec only of the call who suffered the rate change, equal to the SC-rate. The main difference between them is becoming visible when not all STAs are participating in the adaptation process (Figure 3(d)), where codec adaptation is more effective than frame size adaptation, and both
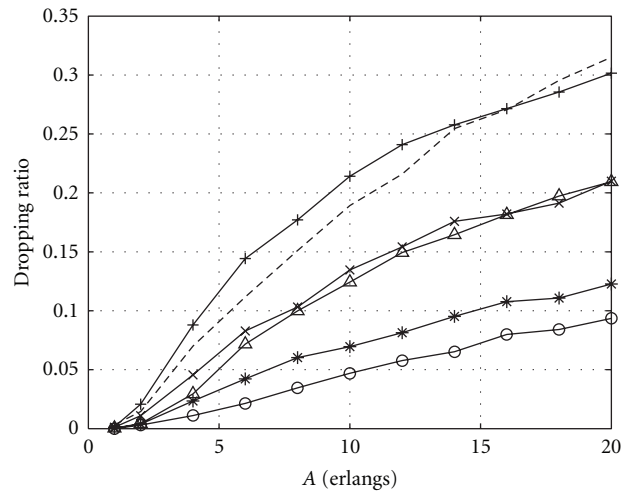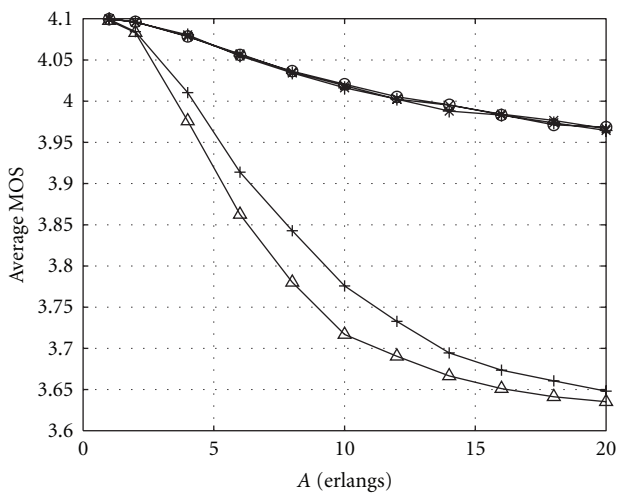
(a) Blocking ratio (100% participation)

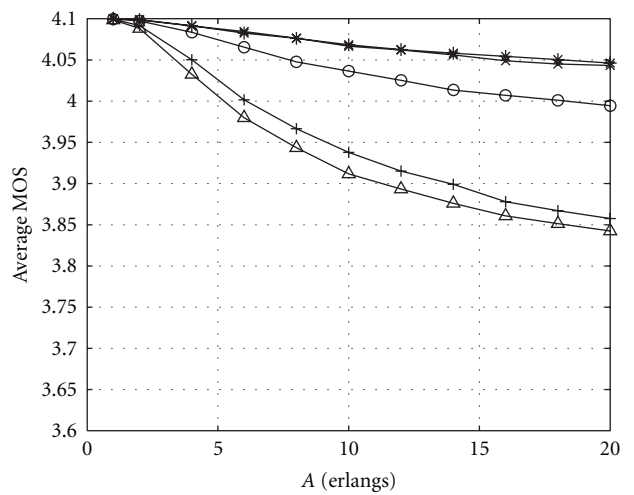(b) Blocking ratio (40% participation)

(c) Dropping ratio (100% participation)

(d) Dropping ratio (40% participation)

(e) Average MOS (100% participation)

(f) Average MOS (40% participation)

FIGURE 3: Performance metrics at different percentages of STAs participation.
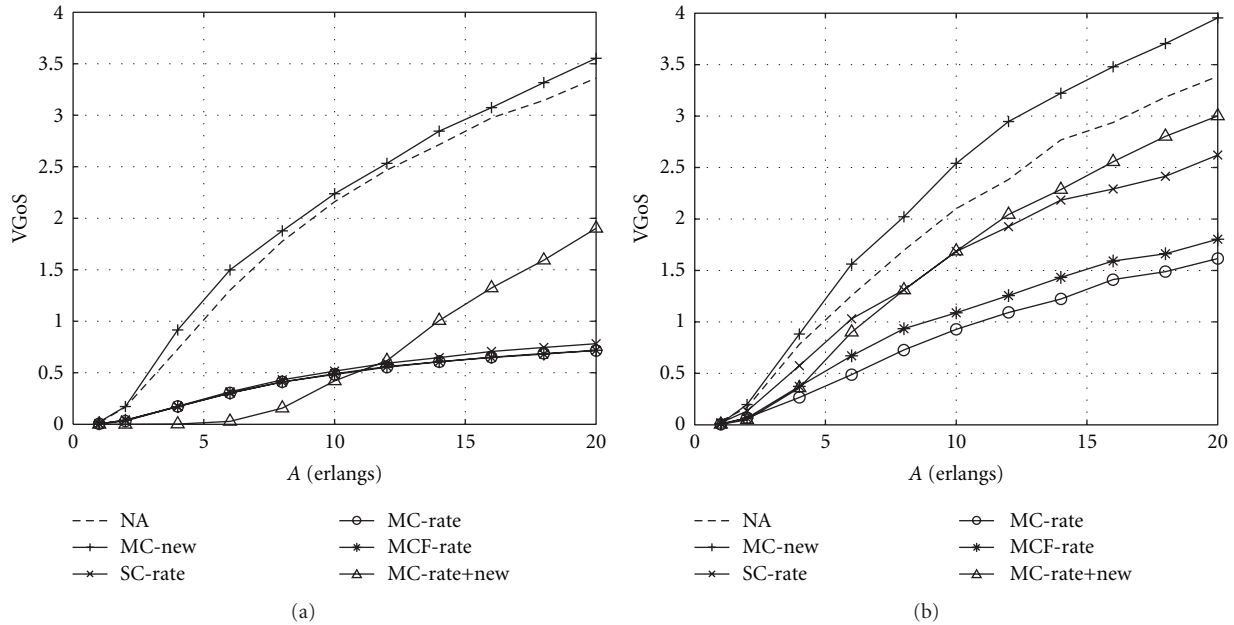
FIGURE 4: VGoS versus offered traffic for (a) 100% node participation, (b) 40% node participation.

are more effective than the single adaptive case. Notice here that in our simulations we used four possible codecs for the adaptation but only two possible packet sizes. This implies that the adaptation capability of MCF-rate is lower than that of the MC-rate. In the general case this is common, since increasing the packet size too much will lead in higher packetization delay, thus it should be avoided.

(4) *Apply Codec Adaptation Both on Arrival of New Calls and at Any Rate Change Causing Congestion (MC-rate+new).* A trade-off between blocking and dropping probabilities is achieved which leads to the highest cell utilization. However, this case shows the worst $E[\text{MOS}]$ values as it allows the highest number of codec adaptations. Notice how the $E[\text{MOS}]$ tends to its minimum value, equal to 3.6, as all active calls are finally using the lowest bandwidth codec (*G.723.1* with MOS equal to 3.6) when traffic load increases. A lower number of nodes participating in the adaptation process results in an increase of both blocking and dropping probabilities as the adaptation capabilities are reduced and the available ones are mainly wasted at the entrance of new calls. As in all other cases, when the network adaptability is reduced, the average MOS increases (less calls active but with higher codecs).

As we can see from the results in Figure 3, regardless of the percentage of the participating nodes in the adaptation process, there is not one solution that clearly outperforms the others in *all* metrics. Although some policies minimize blocking/dropping probability, they have a direct impact on the average MOS of the calls, and vice versa. So whether to choose one policy over the other depends highly on the

parameter that we want to optimize. The trade-off between capacity (in terms of carried traffic) and quality is evident and not always maximizing the capacity is the main scope of a network, especially when dealing with sensitive VoIP calls. On the other hand, having an average MOS of 3.6 could be perfectly acceptable in terms of quality if this can additionally permit to maximize the number of clients accepted.

6.2. Choosing the Optimal Codec Adaptation Strategy. Thus, to decide which solution performs best, we need to consider all metrics together. This is where the VGoS factor can be highly beneficial in order to join all three different metrics into one and facilitate this process. In Figure 4, the VGoS-factor is plotted against the offered traffic A for two values of a number of participating nodes, 100%, and 40%, respectively. The values of blocking, dropping, and $E[\text{MOS}]$ obtained from the previous experiment were combined following (2) in order to calculate the equivalent VGoS value. Remember that, as explained previously, the highest weight in the VGoS calculation is given to the dropping rate parameter.

In average, the three solutions that act on rate changes only (SC-rate, MC-rate and MCF-rate) can give a better overall performance (lower VGoS) under high traffic load when considering all three metrics and when all nodes participate on the adaptation process (Figure 4(a)). On the contrary, MC-rate outperforms clearly the other two and especially compared to the SC-rate (Figure 4(b)). In the case that an STA does not participate in the adaptation procedure, MultiCodec policies can change the codec of another call while SingleCodec cannot. Thus, the multiadaptive option becomes more effective.

In case of the MC-rate+new, an increased number of calls are admitted in the cell using low bitrate codecs. This
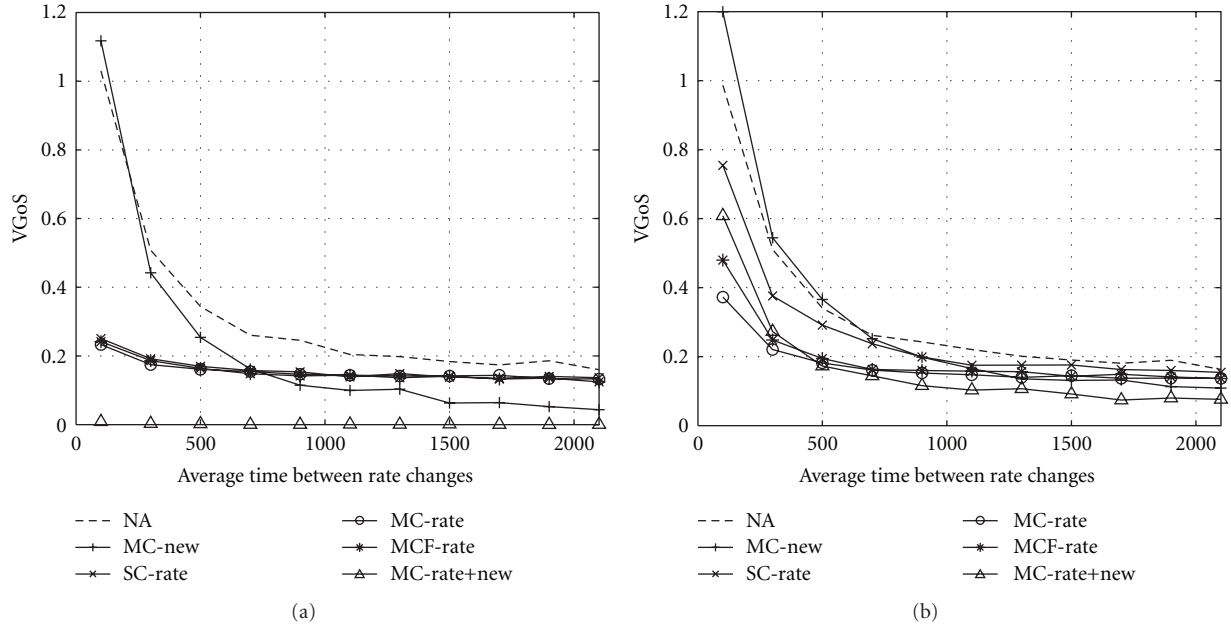
FIGURE 5: VGoS-factor versus frequency of rate changes at 5 erlangs with (a) 100% node participation, (b) 40% node participation.

number of calls is higher than the one obtained by the policies acting on rate changes only, since we apply codec adaptation also at the entrance of new calls (note that this is true regardless of the number of nodes participating on the adaptation process). However, as a consequence, many of the admitted calls are actually dropped afterwards before successfully finishing and especially when the traffic load increases (we waste adaptation capacity at the entrance). This can be depicted at the high VGoS values of the MC-rate+new compared to the "on rate" solutions on high traffic load $A > 12$ erlangs with a 100% of nodes participating or even for lower traffic with a 40% of nodes participating.

Summarizing these observations, a cell under low traffic conditions (note that the threshold depends on the number of participating nodes) has the flexibility to permit codec modifications, so as to try to accept more new calls (overall capacity increase) without heavily impacting the dropping and blocking probabilities. Hence the low VGoS values for the MC-rate+new case in low traffic load. However, under heavier load conditions or under low STAs participation, there is less margin for adaptation and so the best solution in order to keep all three metrics in check is to change codec only as a response to rate changes (the "on rate" usage scenario), and to focus on guaranteeing a low dropping probability for the already accepted calls.

6.3. VGoS-Factor versus Rate Change Frequency. Another basic issue is how the different policies respond to the frequency of transmission rate changes. In Figures 5(a) and 5(b), we plot the VGoS-factor against the time interval between rate changes for a traffic load of 5 erlangs and two different number of participating nodes, 100% and 40%.

As expected, all policies improve their VGoS values when the frequency of rate changes decreases. This is directly

noticeable in the performance achieved by the nonadaptive (NA) solution, which is based on dropping the call that causes congestion after a rate change (less rate changes lead to less call dropping). The same observation is valid for the case of MC-new. Remember that MC-new could obtain the best blocking ratio for adapting the codecs of new calls at the entrance. Additionally, it can now also achieve a low dropping ratio when the frequency of rate changes decreases, thus becoming one of the best policies considering all metrics (Figure 5(a)).

Nevertheless, the MC-rate+new shows the best VGoS under any frequency of rate changes. The only case when it is outperformed is for low number of participating STAs (Figure 5(b)) and high frequency of rate changes (time interval between rate changes less than 200 sec). As explained before, with a high number of rate changes, since the MC-rat+new wastes part of its adaptation capabilities to accept new calls, it then suffers from a higher dropping probability compared with the "on rate" usage. This effect is intensified by the less existing possibilities for codec adaptations in the case of 40% of participating nodes.

7. Conclusions

In this paper, we examine how different solutions based on codec adaptation can be used in a multirate WLAN scenario in order to achieve the best capacity versus quality trade-off. Using voice codec adaptation not only to mitigate the multirate effects but to increase the number of accepted calls in the cell, we can obtain a significant capacity increase, especially in the case of low to medium traffic load.

In order to compare the various policies and scenarios, three different metrics are considered, the blocking rate, dropping rate, and MOS. By uniting all three important

metrics in one new GoS parameter, the VGoS introduced in this paper, a simple yet effective tool for optimizing the adaptation process, is provided. We show that by using the VGoS, we can get valuable conclusions for the performance of the system, the scenario, and the policy that gives the best trade-off between quality and capacity.

Summarizing the main comparison results, we obtain that the use of adaptive policies is always beneficial. However, when (1) the traffic load is low, (2) the frequency of rate changes is low, and (3) the number of participating calls in the adaptation process is high, codec adaptation should be used in both new calls and rate changes, making "MultiCodec on new and rate" (MC-new+rate) policy the best option for these scenarios. Otherwise, it is better to adapt the voice codecs only at rate changes, thus the "MultiCodec on rate" solutions are the ones performing best.

Finally, the effect of the adoption of the codec adaptation solution by the STAs is analyzed. We see that limiting the percentage of the STAs that participate in the process can decrease somewhat the effectiveness of the codec adaptation, especially in high traffic load conditions. However, using codec adaptation, and in particular the multiadaptive solutions, still outperforms the standard nonadaptive case, and the expected terminal evolution will only increase the efficiency of these proposals. Further issues of fairness that arise are left for future study.

References

[1] IEEE Std 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ANSI/IEEE Std 802.11, 1999.

[2] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies*, pp. 836–843, San Francisco, Calif, USA, April 2003.

[3] S. Garg and M. Kappes, "Can I add a VoIP call?" in *Proceedings of IEEE International Conference on Communications (ICC '03)*, Anchorage, Alaska, USA, May 2003.

[4] S. Garg and M. Kappes, "An experimental study of throughput for UDP and VoIP traffic in IEEE 802.11 b networks," in *Proceedings of the Wireless Communications and Networking Conference (WCNC '03)*, 2003.

[5] A. Trad, F. Munir, and H. Afifi, "Capacity evaluation of VoIP in IEEE 802.11e WLAN environment," in *Proceedings of the 3rd IEEE Consumer Communications and Networking Conference (CCNC '06)*, pp. 828–832, January 2006.

[6] B. Bellalta, M. Meo, and M. Oliver, "A BEB-based admission control for VoIP calls in WLAN with coexisting elastic TCP flows," in *Proceedings of the 6th International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking*, vol. 4003 of *Lecture Notes in Computer Science*, pp. 130–141, St. Pertersburg, Russia, 2006.

[7] R. G. Garroppo, S. Giordano, S. Lucetti, and L. Tavanti, "Providing air-time usage fairness in IEEE 802.11 networks with the deficit transmission time (DTT) scheduler," *Wireless Networks*, vol. 13, no. 4, pp. 481–495, 2007.

[8] P. McGovern, S. Murphy, and L. Murphy, "Addressing the link adaptation problem for VoWLAN using codec adaptation,"

[9] J. J. Chen, L. Lee, and Y. C. Tseng, "Integrating SIP and IEEE 802.11e to support handoff and multi-grade QoS for VoIP applications," in *Proceedings of the 2nd ACM international workshop on Quality of service & security for wireless and mobile net- works*, pp. 67–74, 2006.

[10] P. McGovern, S. Murphy, and L. Murphy, "Protection against link adaptation for VoWLAN," in *Proceedings of the 15th IST Mobile and Wireless Communications Summit*, 2006.

[11] A. Sfairopoulou, C. Macián, and B. Bellalta, "VoIP codec adaptation algorithm in multirate 802.11 WLANs: distributed vs. centralized performance comparison," in *Proceedings of the 13th Open European Summer School and IFIP TC6.6 Workshop on Dependable and Adaptable Networks and Services*, vol. 4606, pp. 52–61, July 2007.

[12] A. Sfairopoulou, C. Macián, and B. Bellalta, "QoS adaptation in SIP-based VoIP calls in multi-rate IEEE 802.11 environments," in *Proceedings of the 3rd International Symposium on Wireless Communication Systems (ISWCS '06)*, pp. 475–479, Valencia, Spain, September 2006.

[13] C. Cano, B. Bellalta, and M. Oliver, "Adaptive admission control mechanism for IEEE 802.11e WLANs," in *Proceedings of IEEE Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, Athens, Greece, September 2007.

[14] J. Rosenberg, H. Schulzrinne, G. Camarillo et al., *RFC3261: SIP: Session Initiation Protocol*, Internet RFCs, 2002.

[15] S. Garg and M. Kappes, "Admission Control for VoIP Traffic in IEEE 802.11 Networks," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, pp. 3514–3518, San Francisco, Calif, USA, December 2003.

[16] A. Trad, Q. Ni, and H. Afifi, "Adaptive VoIP transmission over heterogeneous wired/wireless networks," in *Interactive Multimedia and Next Generation Networks*, vol. 3311 of *Lecture Notes in Computer Science*, pp. 25–36, 2004.

[17] T. Kawata and H. Yamada, "Impact of multi-rate VoIP on quality of service in IEEE 802.11e EDCA with link adaptation," in *Proceedings of IEEE International Conference on Communications (ICC '07)*, pp. 392–397, June 2007.

[18] R. G. Garroppo, S. Giordano, S. Lucetti, and E. Valori, "The wireless hierarchical token bucket: a channel aware scheduler for 802.11 networks," in *Proceedings of the 6th IEEE International Symposium on World of Wireless Mobile and Multimedia Networks*, pp. 231–239, 2005.

[19] B. Tebbani and K. Haddadou, "Codec-based adaptive QoS control for VoWLAN with differentiated services," in *Proceedings of the 1st IFIP Wireless Days (WD '08)*, November 2008.

[20] M. F. Tüysüz and H. A. Mantar, "A cross layer QoS algorithm to improve wireless link throughput and voice quality over multi-rate WLANs," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference on ZZZ*, pp. 209–213, 2010.

[21] A. Sfairopoulou, B. Bellalta, and C. Macian, "How to tune VoIP codec selection in WLANs?" *IEEE Communications Letters*, vol. 12, no. 8, pp. 551–553, 2008.

[22] B. Bellalta, C. Macian, A. Sfairopoulou, and C. Cano, "Evaluation of joint admission control and VoIP codec selection policies in generic multirate wireless networks," in *ITC/IEEE NEW2AN'07*, St. Pertersburg, Russia, September 2007.

[23] S. P. Chung and J. S. You, "Call admission control in CDMA cellular networks with grade of service and quality of service dimensioning," *WSEAS Transactions on Communications*, vol. 5, no. 12, pp. 2182–2189, 2006.

[24] ITU-T Rec. G.107, "The E-model, a computational model for use in transmission planning," 2000.

[25] S. A. AlQahtani and A. S. Mahmoud, "Dynamic radio resource allocation for 3G and beyond mobile wireless networks," *Computer Communications*, vol. 30, no. 1, pp. 41–51, 2006.

[26] C. Hoene, H. Karl, and A. Wolisz, "A perceptual quality model intended for adaptive VoIP applications," *International Journal of Communication Systems*, vol. 19, no. 3, pp. 299–316, 2006.

[27] IEEE Std 802.11e, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment: Medium Access Control(MAC) Quality of Service Enhancements," IEEE Std 802.11e, 2005.

[28] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in *Proceedings of IEEE International Conference on Communications*, pp. 196–201, Paris, France, June 2004.

[29] B. Bellalta, M. Oliver, M. Meo, and M. Guerrero, "A simple model of the IEEE 802.11 MAC protocol with heterogeneous traffic flows," in *IEEE Eurocon 2005*, Belgrade, Serbia and Montenegro, November 2005.

[30] NS2, "Network Simulator, release 2.28," February 2005, http://www.isi.edu/nsnam/ns/.

[31] G. Chen, "Component Oriented Simulation Toolkit," 2004, http:// www.cs.rpi.edu/scheng3/.

[32] Cisco Technical Document, "Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation".