On VoIP Capacity Gains in Frame-Based Packet-Switched Wireless Networks

Klaus Sambale and Karsten Klagges
Department of Communication Networks (ComNets)
Faculty 6, RWTH Aachen University, Germany
Email: {Klaus.Sambale|Karsten.Klagges}@comnets.rwth-aachen.de

Abstract—**It is expected that future IMT-Advanced systems will operate packet-switched due to the dominance of data services. However, it is inherent to packet-switched systems to not cope well with voice services as periodic resource usage patterns are not signaled efficiently. In this paper, we present a simple technique to reduce the signaling overhead for resource allocations especially for Voice-over-IP (VoIP) traffic thus increasing the VoIP capacity of the system, considerably. The proposed technique assigns Persistent Resource Allocations (PRAs) per VoIP connection that are valid for a fix number of frames. Resource allocations for succeeding frames have to be signaled by new PRAs. This supersedes any procedures for deallocation of PRAs and associated error handling. Within our work we introduce an analytical model to evaluate the possible VoIP capacity gains. Additionally, we show the potential of statistical multiplexing of VoIP calls. We present simulation results for the packet loss rates resulting from failed resource requests and lack of resources. Both techniques are applicable to all packet-switched systems. We show exemplary results for the WiMAX system.**

I. Introduction

For the evaluation of International Mobile Telecommunications (IMT)-Advanced candidate systems the International Telecommunication Union - Radio Group (ITU-R) Working Party 5D (WP5D) defined some benchmark scenarios [1]. One of these evaluates the VoIP capacity in terms of number of simultaneous calls that can be carried per cell. Hence, it is essential for a candidate system to maximize its VoIP capacity. As this performance indicator is also known to be of great interest by network operators high performance in this field will strengthen the position of a system in the market.

The Media Access Control (MAC) protocols of many standards originating from the computer world are optimized for packet-switched data traffic, not considering the special traffic characteristics of voice services. In contrast, Global System for Mobile communications (GSM) and Universal Mobile Telecommunications System (UMTS) have been initially developed for circuit-switched voice services, only. As the demand for data services increased these systems have been enhanced to carry packet-switched services. But the support of these services is still not optimal. As data services are expected to dominate the traffic in future wireless networks IMT-Advanced systems will operate packet-switched. Nevertheless, some techniques developed for systems such as GSM can be transferred to packet-switched networks to improve their performance for voice services.

One of the most promising techniques to be transferred is the (timely limited) fixed allocation of resources to certain Mobile Stations (MSs) within subsequent frames. Thus, the repeatedly signaling of a periodic resource usgae pattern can be avoided. The signaling overhead can be reduced and the overall capacity increased. There are already some proposals, e.g. [2], to implement this technique. In the current draft of the next revision of the Institute of Electrical and Electronics Engineers (IEEE) 802.16 standard [3] a complex mechanism for the persistent allocation of resources is proposed. But the advantages of reducing the signaling overhead are diminished through additional signaling necessary for allocation, re-allocation and deallocation of persistent allocations. Additionally, sophisticated error handling procedures have to be established to avoid interference when signaling messages get lost.

In this paper, we propose a simple technique for PRAs that greatly increases the VoIP capacity of packet-switched systems and that supersedes any error handling procedures. Furthermore, we present possible VoIP capacity gains through statistical multiplexing. Both techniques are applicable to all packet-switched systems. Within our work we show exemplary results for the IEEE 802.16 standards family often also referred as Wireless Interoperability for Microwave Access (WiMAX).

The paper is structured as follows: In Section II we introduce our concept for persistent resource allocation. Then, in Section III we present the analytical model that is the basis for the results discussed in Section IV. The conclusions are summarized in Section V.

II. Concept

The traffic of voice services is characterized by constant VoIP packet sizes and Inter Arrival Times (IATs) during talk spurts and no or very rare transmissions for comfort noise data during pauses. In current frame-based packet-switched networks the necessary resources for voice services are periodically assigned to the MSs and signaled at the start of each frame as indicated in the upper part of Fig. 1.

As mentioned above many proposals imply an extension of the validity of signaled resource allocation to a certain number of periodic repetitions. As the IAT and size of VoIP packets are fixed and known the period length and the amount of resources necessary per repetition are fixed, too. We propose

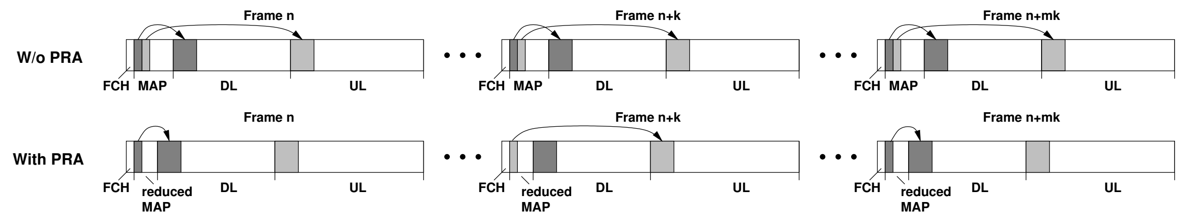

Fig. 1. On the upper part the signaling of resource allocations as applied in current communication networks is shown: The allocation of all resources of a frame are signaled at the frame start. On the lower part the signaling of resource allocations of PRA enabled communication networks is depicted: The resource allocations for connections with periodic resource requests with period length k are valid for m frames.

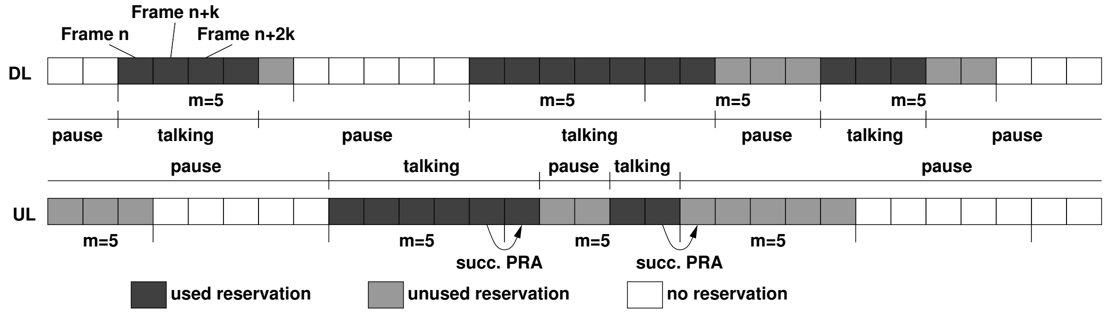

Fig. 2. Resource usage of a single PRA enabled VoIP connection

to set the validity of a PRA to a fix number of repetitions as depicted in the lower part of Fig. 1: the resource allocations signaled in frame n for two different voice connections, one in Up-Link (UL) and one in Down-Link (DL) direction, remain valid in frame $n + k$ up to frame $n + (m - 1)k$. If the voice connections are ongoing and hence need further resources as exemplary shown on the lower right side of Fig. 1 the resource allocations have to be signaled in frame $n+mk$, again. We call m Time-To-Live (TTL) as it limits the life time of a periodic resource allocation. The parameter k matches the period length of packet arrival in number of frames. This parameter is fixed for a certain frame length and VoIP codec. A typical value for k is 4 at a frame length of 5 ms and an IAT of 20 ms for VoIP packets. The resources of a frame envisaged for signaling resource allocations that are saved by applying the PRA scheme can be allocated to other VoIP connections. Thus, the number of VoIP connections that can be carried per frame can be increased.

Transmitting the two parameters m and k whenever a resource allocation is signaled would induce additional signaling overhead. As k is fixed for a certain VoIP connection and m can be chosen to be fixed, too, we propose to negotiate both parameters once during the service flow setup of the voice connection. Furthermore, we assume that each VoIP packet is transmitted in a new burst on the Physical Layer (PHY) and that no concatenation of VoIP packets in DL direction is applied for the following reason: In general, MSs have to decode PHY bursts completely. But it is a waste of energy for battery powered devices if not all data within a PHY burst are addressed to them. Hence, transmitting each VoIP packet in a new PHY burst increases the average talk times of the MSs.

For real-time interactive services like VoIP traffic some standards such as the IEEE 802.16 recommend the use of Unsolicited Grant Service (UGS). For all service flows that use UGS the Base Station (BS) periodically allocates DL and UL resources to maintain a minimum reserved traffic rate negotiated at service flow setup. The resources are allocated independently of the state the VoIP connection is in (“talking” or “pause”). In UL direction this supersedes bandwidth requests. The PRA technique together with UGS already increases the VoIP capacity. A further considerable increase in VoIP capacity can be expected through statistical multiplexing of VoIP connections as typically only one partner in a voice call is talking at a time. Hence, most of the time only UL or only DL traffic has to be carried per VoIP connection. To allow for statistical multiplexing UGS may not be used. This raises a challenge for the UL direction of current packet-switched networks: If a MS wants to transmit VoIP data in UL direction it requests UL bandwidth via contention slots. If it gets assigned an UL resource it transmits the VoIP packet. As the period length between succeeding VoIP packets is usually a multiple of the frame duration, the MS does not request further bandwidth piggy-backed as no further packets are waiting to be sent. Hence, for each new UL VoIP packet a new bandwidth request has to be sent via a contention slot. If multiple MSs have VoIP connections this results in collisions on the contention access slots and thus no MS actually gets any resources assigned. The PRA technique greatly reduces the number of bandwidth requests that have to be transmitted via contention slots reducing the collision probability also in heavy loaded cells considerably.

In Fig. 2 the resource usage of a single VoIP connection applying the PRA concept is shown for $m = 5$. In DL direction the first resource of a PRA phase is always used. Otherwise no PRA would be signaled for that connection. The remaining resources of that PRA phase may be used or not. The same applies in UL direction for the first PRA after a long pause without a valid PRA. But to allow for statistical multiplexing of VoIP connections and to avoid collisions on the contention slots the BS automatically assigns succeeding PRAs in UL direction if the last allocated resource of a valid PRA phase is used for the transmission of a VoIP packet. As a result succeeding PRA phases in UL direction can be completely unused as exemplarily shown on the lower right side in Fig. 2. The differences in the usage patterns of the PRA phases have been considered for the analytical model presented in the following section.

By using contention access for signaling resource requests and statistical multiplexing of VoIP connections additional sources for packet errors occur: If several consecutive contention accesses of a MS result in collisions VoIP packet may get out-dated and thus dropped. Additionally, if due to statistical effects more MSs request resources than available, VoIP packets may also get out-dated and dropped. Hence, the total Packet Error Rate (PER) results from transmission errors, failed resource requests and lack of resources.

III. Analytical Model

In [1], the ITU-R WP5D defines a simple 2-state VoIP traffic model as shown in Fig. 3 that shall be used for the evaluation of IMT-Advanced candidate systems. The according model parameters are listed in Tab. I.

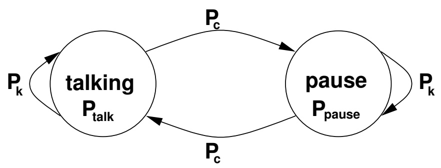

Fig. 3. Markov chain of the VoIP model defined by the ITU-R

TABLE I
VoIP model parameters

Model parameter	Value
Codec	RTP AMR 12.2
Encoder frame length	20 ms
$P_{state} = P_{pause} = P_{talk}$	0.5
P_c	0.99
$P_k = 1 - P_c$	0.01

As mentioned earlier we evaluate the proposed techniques exemplary for the IEEE802.16 standard [3]. Hence, the following calculations base upon a typical parameter set for IEEE802.16 systems. The size of a PHY Protocol Data Unit (PDU) carrying a single VoIP packet is shown in Tab. II. The Payload Header Suppression (PHS) is an optional technique to reduce the PHY PDU size by suppressing that parts of higher layer protocol headers that remain identically on a per packet basis. Only those parts of the headers that are continuously changing are transmitted within each PHY PDU. The constant parts of the header are transmitted once at service flow setup and are referenced by a so called Payload Header Suppression Index (PHSI) later on.

TABLE II
Size of a Physical Layer PDU carrying a single VoIP packet

	W/o PHS [bit]	**With PHS [bit]**
VoIP PDU size	244	244
RTP header	96	48
UDP header	64	0
IP header	160	0
IEEE 802.16 MAC header	48	48
PHSI	-	8
PHY PDU size	612	348

Parts of the IEEE 802.16 MAC frame are reserved for special purposes and hence cannot be used for the transmission of MAP Information Elements (IEs) or DL/UL traffic. In Tab. III the number of symbols for the different MAC frame phases used in our analytical evaluation are listed. We assume a frame length of $t_{frame} = 5\,\text{ms}$ and a system bandwidth of 20 MHz. The analytical evaluation is exemplary performed for the Orthogonal Frequency Division Multiplex (OFDM) PHY. Hence, a MAC frame consists of $N_{frame} = 360$ OFDM symbols. The number of symbols available for the transmission of MAP IEs and DL/UL data bursts calculates to:

$$\begin{aligned} N_{ava} &= N_{frame} - N_{pre} - N_{FCH} - N_{RTG} \\ &\quad - N_{TTG} - N_{rang} - N_{bw} \end{aligned} \tag{1}$$

TABLE III
Typical numbers of OFDM symbols for the different MAC frame phases

Frame phase	**OFDM symbols**
Preamble (N_{pre})	2
FCH (N_{FCH})	1
Receive-Turnaround-Gap (N_{RTG})	4
Transmit-Turnaround-Gap (N_{TTG})	4
Ranging (N_{rang})	20
Bandwidth-Request (N_{bw})	30

To estimate the number of MSs that can be served by a single cell the average PHY burst size for a VoIP packet has to be calculated. For the analytical evaluation we assume a single cell scenario with omni-directional antennas at the BS and at the MSs and free space propagation. Considering the minimum Signal-to-Interference+Noise-Ratio (SINR) requirements for the different PHY modes specified in [3] and assuming that always the best PHY mode is selected the percentage of the area covered by the different PHY modes can be calculated. The results are shown in Tab. IV. Additionally, the number of OFDM symbols necessary to transmit a PHY burst carrying a VoIP packet are listed in that table. The numbers are estimated by dividing the number of bits per PHY burst by the number of bits that can be carried per OFDM symbol for the different PHY modes and rounding up the result.

TABLE IV
USAGE OF PHY MODES AND NUMBER OF OFDM SYMBOLS PER VOIP PACKET IN A FREE SPACE SCENARIO

Modulation	Coding Rate	Coverage [%]	Bit per Symbol	Symbols w/o PHS	Symbols with PHS
BPSK	1/2	39.40	96	7	4
QPSK	1/2	26.52	192	4	2
QPSK	3/4	17.00	288	3	2
16QAM	1/2	9.46	384	2	1
16QAM	3/4	4.59	576	2	1
64QAM	2/3	1.12	768	1	1
64QAM	3/4	1.92	864	1	1

The average number of symbols $N_{avg,sym}$ per PHY burst calculates to:

$$N_{sym,avg} = \frac{\sum_{\text{PHY modes}} \frac{R(\text{PHY mode})}{100} \cdot N(\text{PHY mode})}{g_{SDMA}} \quad (2)$$

where $R(\text{PHY mode})$ denotes the percentage of the coverage area by that PHY mode and $N(\text{PHY mode})$ the number of symbols necessary to transmit the PHY burst. The parameter g_{SDMA} represents the capacity gain that can be reached if optionally Space Division Multiple Access (SDMA) is applied. For that case we assume that up to four MSs at different locations within the cell can be served the same time. The results in [4] show that this technique leads to a maximum performance gain of about 3.2 under conditions comparable to the scenario presented herein (single cell scenario, free space propagation, uniformly distributed MSs, fixed Equivalent Isotropic Radiated Power (EIRP)). The results are listed in Tab. V. Furthermore, the DL (N_{DL-MAP})/UL (N_{UL-MAP}) MAP sizes are shown in the table. The numbers are mapped to numbers of OFDM symbols assuming that the Binary Phase Shift Keying (BPSK) 1/2 PHY mode is used for the transmission of the MAPs. The optional SDMA mode cannot be applied during the transmission of the DL/UL MAP data as stated in [4]. The according size of the UL preamble N_{UL-pre} is shown in the last row of the table.

TABLE V
AVERAGE PHY BURST SIZES FOR VOIP PACKETS

	W/o SDMA		With SDMA	
	W/o PHS	With PHS	W/o PHS	With PHS
N_{avg} [symbols]	4.55	2.62	1.41	0.81
g_{SDMA}	1		3.22	
DL-MAP	32bit = 0.33 symbols		48bit = 0.5 symbols	
UL-MAP	48bit = 0.5 symbols			
UL burst preamble	1 symbol		0.31 symbols	

VoIP connections generate a symmetric traffic load. Hence, on average the same number of PHY bursts is transmitted in DL direction as in UL direction. Therefore, the average PHY burst size N_{data} can be calculated to be independent of the transmission direction without loosing exactness of results:

$$N_{data} = N_{sym,avg} + \frac{1}{2} N_{UL-pre} \quad (3)$$

The same applies to the MAP sizes:

$$N_{MAP} = \frac{1}{2}(N_{UL-MAP} + N_{DL-MAP}) \quad (4)$$

Considering all assumptions made so far the VoIP capacity applying UGS can be estimated by the following formula:

$$N_S = \frac{N_{ava}}{2\left(N_{data} + \frac{N_{MAP}}{m}\right)} \cdot \frac{t_{IAT}}{t_{frame}} \quad (5)$$

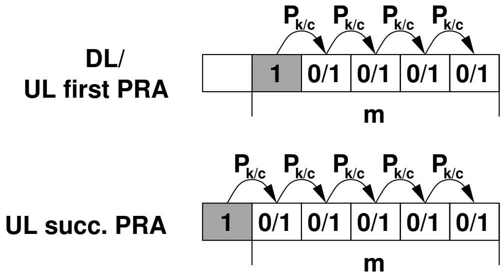

Fig. 4. Resource usage patterns and transition probabilities

As already mentioned in Section II further capacity gains can be reached by statistical multiplexing. For the estimation of the VoIP capacity when applying statistical multiplexing the resources of a valid PRA that are not used by the MS have to be considered as overhead as they cannot be reused by other MSs. This diminishes the capacity gain. In the following we derive the formula to calculate this overhead.

In Fig. 4 all combinations of valid PRAs are shown. In DL direction and for the first PRA after a long pause in UL direction the first resource of a PRA is always used indicated by "1". The remaining resources of that PRA may be used either ("1") or not ("0"). Unlike, for succeeding PRAs in UL direction already the first resource may not be used. The allocation of that PRA results from the transmission of a PHY burst in the last allocated resource of the previous PRA. The average overhead arising from unused allocated resources can be estimated by adding the probabilities of all possible resource usage patterns multiplied by the according number of "0s" for each usage pattern.

Equ. 6 taken from [5] calculates the number of runs of "1"s and "0"s for the binary representation of an integer value. Subtracting 1 from the result equals the number of changes between "1"s and "0"s for the binary representation of that integer value. Using this value and the state change probabilities of the VoIP model the probability of a certain PRA resource usage pattern can be derived.

$$a(2^k + i) = a(2^k - i + 1) + 1 \text{ for } k \geq 0 \text{ and } 0 < i \leq 2^k \quad (6)$$

To calculate the number of zeros in a binary representation of an integer value formula Equ. 7 is used. The formula is

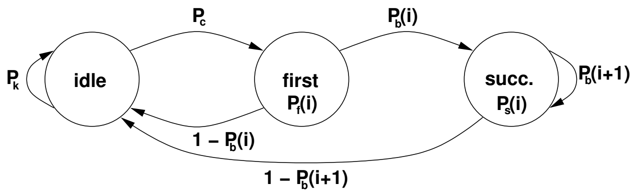

Fig. 5. Markov chain on that the calculation of the ratios between first and succeeding PRAs in UL direction is based.

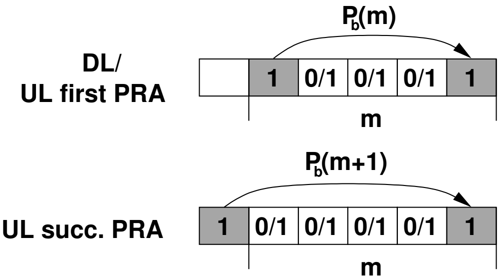

Fig. 6. The state change probabilities of the Markov model shown in Fig. 5 reflect the probability that the last resource of a valid PRA is used. For the first PRA in UL direction after a long pause (shown in the upper part) this probability is higher than for succeeding PRAs (shown in the lower part).

proposed in [6].

$$b(n) = \begin{cases} 0 & \text{for } n < 1 \\ b(\lfloor \frac{n}{2} \rfloor) + 1 - n \mod 2 & \text{else} \end{cases} \tag{7}$$

Using Equ. 6 and Equ. 7 the average overhead $N_o(m)$ can be determined to:

$$N_o(m) = \sum_{i=0}^{2^{m-1}} \left(P_c^{a(i+2^{m-1})-1} b(i+2^{m-1}) P_k^{m-a(i+2^{m-1})-2} \right) \tag{8}$$

As already mentioned the resource usage patterns of the first and succeeding PRAs in UL direction are different. Hence, to determine the average PRA overhead for a given m in UL direction the ratio of first to succeeding PRAs has to be determined. The Markov model shown in Fig. 5 reflects this situation. If an MS is in state "idle" it remains there with probability P_k. With probability P_c it generates UL traffic. Then, it changes to state "first" and gets a first PRA. The probability $P_b(i)$ that a succeeding PRA is requested can be derived by Equ. 9 and Equ. 10. As shown in the upper part of Fig. 6 $P_b(i)$ is the probability that the last resource of the first UL PRA is used and hence a succeeding PRA is allocated. In this case the MS changes to state "succ.". With probability $1 - P_b(i)$ the MS returns to state "idle". Further succeeding PRAs are only requested if the last resource of the current PRA is used. Hence, the probability of staying in state "succ." is $P_b(i+1)$. The subfigure on the lower part of Fig. 6 illustrates this. With probability $1 - P_b(i+1)$ the MS does not request further UL resources and changes back to state "idle".

$$P_b(i) = \begin{cases} 1 & \text{for } i = 1 \\ P_b(i-1)P_k + P_i(i-1)P_c & \text{else} \end{cases} \tag{9}$$

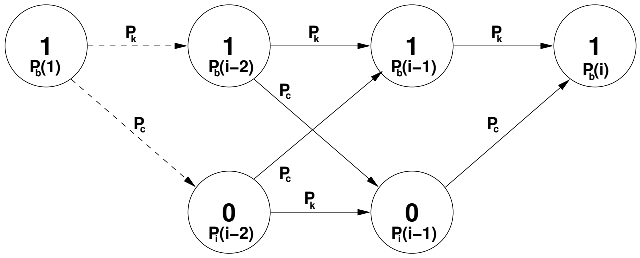

Fig. 7. The probability $P_b(i)$ that the last resource of a valid PRA is used can be derived by this Markov chain. Eqn. 9 and Eqn. 10 describe this model.

$$P_i(i) = \begin{cases} 0 & \text{for } i = 1 \\ P_i(i-1)P_k + P_b(i-1)P_c & \text{else} \end{cases} \tag{10}$$

For the above mentioned Markov model the steady state probabilities of being in state "first" $P_f(i)$ and state "succ." $P_s(i)$ can be calculated. But for the estimation of the average PRA overhead in UL direction only the ratios between first and succeeding PRAs are relevant. These ratios are determined by the following two formulas:

$$R_f(i) = \frac{P_f(i)}{P_f(i) + P_s(i)} = \frac{1 - P_b(i+1)}{1 + P_b(i) - P_b(i+1)} \tag{11}$$

$$R_s(i) = \frac{P_s(i)}{P_f(i) + P_s(i)} = \frac{P_b(i)}{1 + P_b(i) - P_b(i+1)} \tag{12}$$

Then, the mean overhead of PRAs $N_{oh}(m)$ independent of the transmission direction calculates to:

$$N_{oh}(m) = \frac{1}{2}\big((1 + R_f(m)) \cdot N_o(m) + R_s(m) \cdot N_o(m+1)\big) \tag{13}$$

Unlike, the overhead resulting from PRAs that is increasing with m the overhead resulting from the MAP signaling decreases for higher m. The total overhead considering both effects $N_{oh,tot}$ can be derived as follows:

$$N_{oh,tot}(m) = \frac{N_{oh}(m) + N_{MAP}}{N_{oh}(m) + N_{MAP} + mN_{data}} \tag{14}$$

The total overhead per PRA together with the average number of OFDM symbols per PHY burst can now be mapped to an estimated cell capacity $N_{s,cap}$ when considering the voice activity factor $P_{state} = P_{talk} = 0.5$:

$$N_{S,cap}(m) = \frac{(1 - N_{oh,tot}(m)) \cdot N_{ava}}{N_{data}} \cdot \frac{t_{IAT}}{t_{frame}} \tag{15}$$

The number of simultaneous calls $N_{S,cap}(m)$ is a maximal number. Due to statistical effects it is possible that MSs request more UL or DL resources than available. To cope with this situation a number of guard resources should be reserved. Thus, the blocking probability can be reduced to a reasonable degree. The trade-off between the number of guard resources and the Quality of Service (QoS) the VoIP users experience is not modeled. Instead, we show simulation results as the analytical model for the blocking probabilities is still under development. Furthermore, we present some simulation results

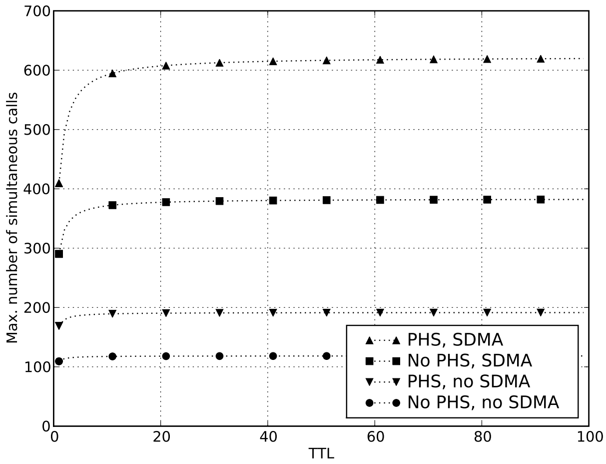

Fig. 8. Max. number of simultaneous calls carried vs. TTL when using the UGS mode for VoIP connections

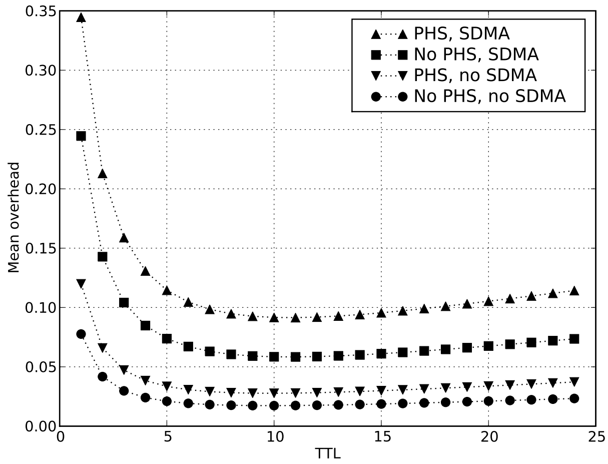

Fig. 9. Mean overhead resulting from signaling of PRAs and unused resources of valid PRAs vs. TTL

indicating the effects of collisions during the contention access phase for UL bandwidth requests.

IV. Results

Fig. 8 shows the maximum number of simultaneous calls vs. TTL applying UGS. As expected the number of calls that can be carried increases with the TTL value but converges asymptotically towards a fixed value that is different for each combination of the optional techniques PHS and SDMA. The overhead resulting from the MAP signaling decreases reciprocally proportionally with increasing TTL values. Overhead arising from PRAs is not considered for UGS as the according resources for each VoIP connection remain allocated no matter if they are really used or not. The capacity gain compared to $TTL = 1$ without using PHS and SDMA converges towards 8 %, when using PHS towards 13 %, when using SDMA towards 30 % and when using both optional techniques towards 49 %. The increasing gains when applying the optional techniques result from the decreasing ratio between the average size of a PHY data burst N_{data} and the average size of a MAP IE N_{MAP}. The smaller this ratio is the higher is the gain.

The graphs in Fig. 9 present the total overhead $N_{oh,tot}$ vs. the TTL value for the case that statistical multiplexing is applied. These graphs reflect both effects that account for overhead. For values up to $TTL = 10$ the overhead resulting from the MAP IE signaling dominates. Hence, the overhead decreases with increasing TTL values. But for values higher than $TTL = 10$ the overhead resulting from unused PRAs starts to dominate the total overhead. Hence, the overhead starts to increase, again. Thus, the optimal TTL value to minimize the total overhead is 10 independent of the application of the optional techniques SDMA and PHS.

The corresponding VoIP capacity graphs are shown in Fig. 10. Again, the number of calls that can be carried per cell are plotted vs. TTL. The capacity gain for the optimal value

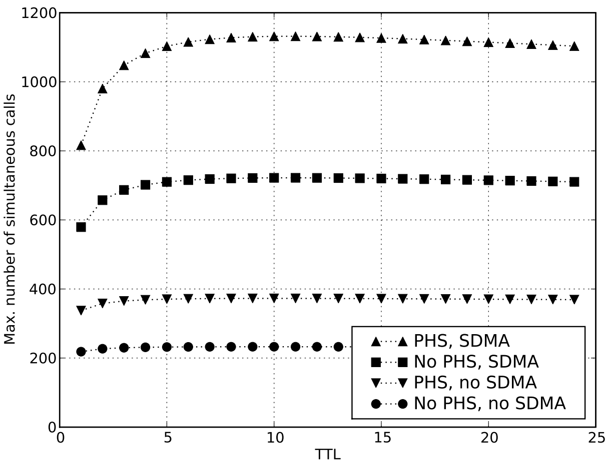

Fig. 10. Max. number of simultaneous calls carried vs. TTL when using statistical multiplexing of VoIP connections

$TTL = 10$ compared to $TTL = 1$ is about 6.5 % without using PHS and SDMA, about 10.5 % when using PHS, about 24.6 % when using SDMA, and about 38.6 % when using PHS and SDMA.

In the following we show preliminary simulation results for the collision probability of the contention access and the corresponding PER as well as for the resource blocking probability and the resulting PER. For the simulations we assumed a frame length of 5 ms. We used the VoIP model described in the previous section. If a new VoIP packet was generated before the previous one had been transmitted, the previous one was dropped. Hence, after four consecutive collisions during contention access or when in four consecutive frames no resources were available for the MS the VoIP packet was dropped.

Figure 11 shows the collision probability for a single

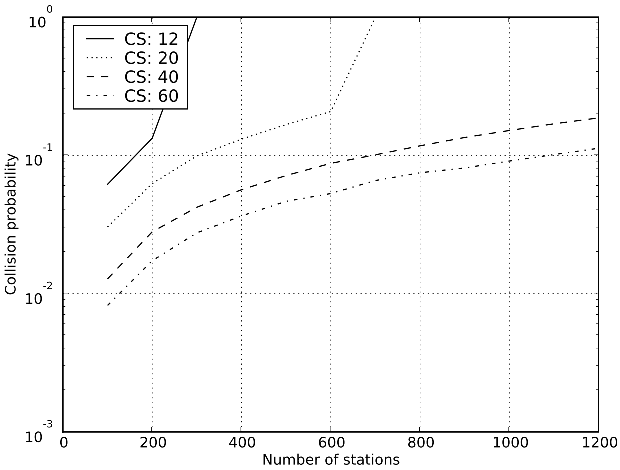

Fig. 11. Collision probability vs. number of MSs for different numbers of contention slots

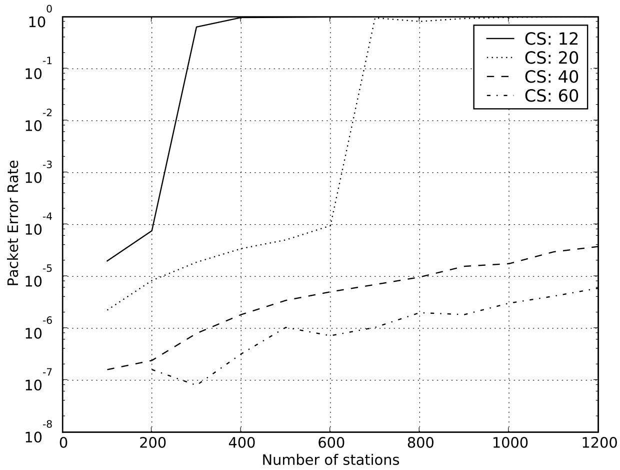

Fig. 12. PER vs. number of MSs for different numbers of contention slots

transmission of a resource request for different numbers of contention slots. The number of contention slots refers to 20 ms. Therefore, if the number of contention slots equals 12 there are 3 contention slots per 5 ms frame. As shown in Figure 10 the maximum number of simultaneous calls when using PHS and SDMA is about 1100. For that number of calls even for 60 contention slots per 20 ms the collision probability is above 0.1. But, as mentioned above only after 4 unsuccessful contention accesses a VoIP is dropped. Therefore, the actual packet loss rate is much lower. Furthermore, contention access has to be used only at the start of a talk spurt as described in Section II. So, if a single resource request is transmitted successfully at the start of a talk spurt for all succeeding VoIP packets no further resource requests have to be transmitted and hence no contention access is necessary. This considerably reduces the PER resulting from the contention access. The corresponding PER is shown in Figure 12. For about 1100 MSs the PER is below 10^{-4} even for 40 contention slots.

The advantage of using contention based access for signaling the resource requests is the small amount of resources that are necessary for the signaling compared to the real-time Polling Service (rtPS). Assuming that 1000 VoIP connections are active about 500 are in state “pause”. That means 500 MSs have to be polled regularly to check if they have any resource needs. To allow a MS to answer a poll at least 2 OFDM symbols have to be allocated to the station (one for the UL preamble and one for the transmission of the bandwidth request). When using SDMA the BS has to assign $1000/3.2 = 312.5$ OFDM symbols per 20 ms or about 80 symbols per frame to MSs for polling. Using contention access only 10 contention slots per 5 ms frame are necessary to keep the PER reasonably low. 10 contention slots equals 20 OFDM symbols. The saved resources (60 OFDM symbols per 5 ms frame or 240 OFDM symbols per 20 ms) corresponds to about 200 VoIP connections. So, contention based signaling

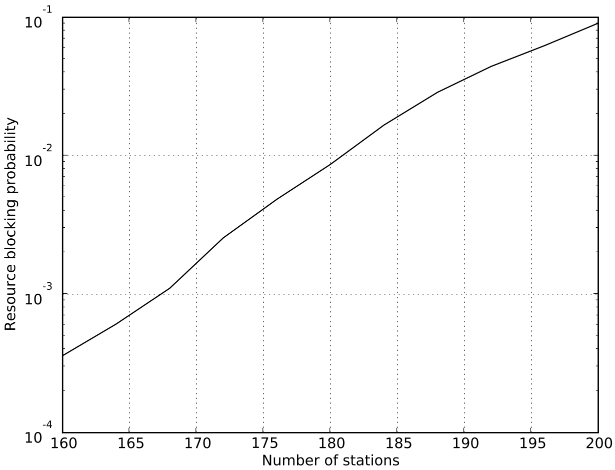

Fig. 13. Blocking probability vs. number of MSs when 100 resources are available

increases the VoIP capacity considerably compared to rtPS while only marginally increasing the PER.

In Figures 13 and 14 the resource blocking probabilities vs. the number of VoIP connections are shown for the case that 100 and 550 resources respectively are available. Per 5 ms frame there are only a fourth of the resources available that means 25 and about 138 respectively. The blocking probability indicates the probability that a resource request cannot be met within the current frame. But, only if also in the succeeding three frames no resources are available the VoIP packet is actually dropped because a new one arrives. Therefore, the PER resulting from the statistical multiplexing is much smaller as shown in Figures 15 and 16.

In the VoIP model used in the evaluation the voice activity factor is 0.5. Hence, theoretically about 200 and 1100 VoIP connections respectively could be active the same time. Due

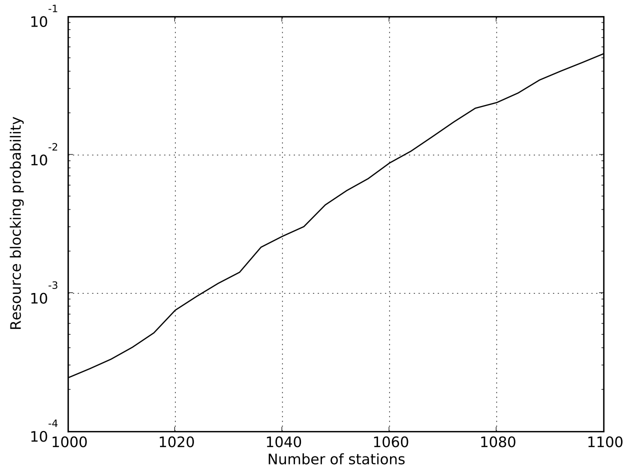

Fig. 14. Blocking probability vs. number of MSs when 550 resources are available

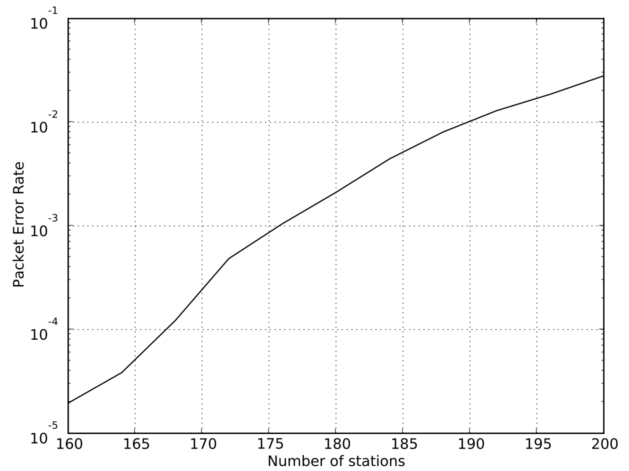

Fig. 15. PER vs. number of MSs when 100 resources are available

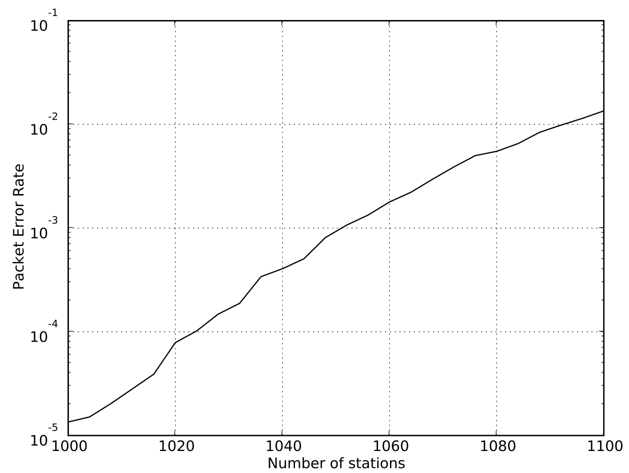

Fig. 16. PER vs. number of MSs when 550 resources are available

to statistical effects for this number of connections the PER is quite high (about 0.03 for 200 and 0.015 for 1100). But when in the case that 100 resources are available only about 175 VoIP connections are active the PER drops below 0.001. The same applies when in case of 550 available resources only about 1050 VoIP connections are active.

Then, the total Packet Error Rate P_{tot} calculates to:

$$P_{tot} = 1 - (1 - P_{trans})(1 - P_{CA})(1 - P_{RA}) \qquad (16)$$

where P_{trans} denotes the PER for the transmission, P_{CA} the PER due to the contention access for signaling bandwidth requests and P_{RA} the PER due to failed resource allocations. According to [1] P_{tot} of VoIP traffic may not exceed 0.02. If we chose 40 contention slots per $20\,\text{ms}$ for signaling resource allocations P_{CA} is always below 0.0001. Furthermore, if we chose the number of VoIP connections in such a way that the corresponding P_{RA} always is below 0.001, P_{trans} has to be below 0.019 to guarantee that P_{tot} stays below 0.02. In case contention access and statistical multiplexing are not used P_{trans} could be equal to 0.02. That means using both techniques requires a reduction of the maximum P_{trans} by 5%, only.

V. Conclusions

The analytical results show that a considerable gain in VoIP capacity can be achieved through the Persistent Resource Allocation (PRA) technique presented in this paper. The technique is simple to implement and supersedes any error handling procedures required by other PRA techniques. Additionally, simulation results indicate that the PRA technique greatly reduces the number of resource requests for UL transmissions sent during the contention access phase thus avoiding collisions. Furthermore, if the maximum number of simultaneous VoIP connections is reasonably limited statistical multiplexing of VoIP connections is feasible without violating the requirements set by ITU-R.

Acknowledgment

This work has been performed in the framework of the FP7 project ROCKET IST-215282 STP, which is funded by the European Community. We would like to acknowledge the contributions of our colleagues from ROCKET Consortium (http://www.ict-rocket.eu).

References

[1] "Guidelines for evaluation of radio interface technologies for IMT-Advanced," ITU-R, Tech. Rep. ITU-R M.2135, Nov. 2008.

[2] S. Kalyanasundaram, A. Bedekar, S. Xu, and H. Xu, "Resource Allocation Scheme for 802.16m," Motorola, Tech. Rep. C802.16m-07/258, Nov. 2007.

[3] *802.16 Rev2 Part 16: Air Interface for Broadband Wireless Access Systems*, IEEE Std. 802.16Rev2, Rev. P802.16Rev2/D9, Jan. 2009.

[4] Hoymann, C., "IEEE 802.16 Metropolitan Area Network with SDMA Enhancement," Ph.D. dissertation, Aachen University, Lehrstuhl für Kommunikationsnetze, Jul 2008. [Online]. Available: http://www.comnets.rwth-aachen.de

[5] The On-Line Encyclopedia of Integer Sequences. [Online]. Available: http://www.research.att.com/~njas/sequences/A005811

[6] The On-Line Encyclopedia of Integer Sequences. [Online]. Available: http://www.research.att.com/~njas/sequences/A080791