EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS
Eur. Trans. Telecomms. 2010; **21**:750–759
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/ett.1455

Persistent resource allocations for VoIP traffic in packet-switching mobile cellular networks[†]

K. Sambale*, K. Klagges and R. Rezai Rad

ComNets Research Group, Prof. Dr. Bernhard Walke, Faculty 6, RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany

SUMMARY

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 signalled efficiently. In this paper, we present a simple technique to reduce the signalling 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 allocation (PRA) per VoIP connection that are valid for a fix number of frames. Resource allocations for succeeding frames have to be signalled 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 results of event-driven simulations to validate the analysis. The PRA technique is applicable to all packet-switched systems. We show exemplary results for the WiMAX system.

1. INTRODUCTION

In current mobile communication systems according to the Global System for Mobile communication (GSM) or Universal Mobile Telecommunications System (UMTS) standard voice calls are transmitted mainly circuit-switched. By contrast, in the International Mobile Telecommunications—Advanced (IMT-Advanced) candidate systems such as Long-Term Evolution (LTE)/Long-Term Evolution-Advanced (LTE-A) or Wireless Interoperability for Microwave Access (WiMAX) voice calls are expected to be transmitted mainly packet-switched as voice-over-IP (VoIP). So, mobile cellular networks are undergoing the same development like the plain old telephony service (POTS): To simplify the network architecture and maintenance all services including voice calls, media streaming, web surfing and file download are provided by a single integrated all-internet protocol (IP) network.

For the evaluation of IMT-Advanced candidate systems the International Telecommunication Union—Radio Group (ITU-R) working party 5D (WP5D) defined several benchmark scenarios [3]. One of those assesses the VoIP capacity in terms of number of simultaneous calls that can be carried per cell. The candidate systems have to comply with minimum requirements set by the ITU-R to become an IMT-Advanced system. Therefore, it is essential for a candidate system, amongst others, 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 its position 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 traffic characteristics specific for voice services. In contrast, GSM and UMTS have been initially developed for circuit-switched voice services, only [2]. As the demand for data

* Correspondence to: K. Sambale, ComNets Research Group, Prof. Dr. Bernhard Walke, Faculty 6, RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany. E-mail: klaus.sambale@comnets.rwth-aachen.de

[†] A previous version of this paper was presented in the 15th European Wireless Conference (EW 2009), Aalborg, Denmark [1].

Received 30 November 2009
Revised 18 June 2010
Accepted 6 September 2010

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 re-used in packet-switched networks to improve their performance for voice services.

One of the most promising techniques to be transferred is the persistent allocation of resources to mobile stations (MSs) within subsequent MAC frames. Thus, the repeated signalling of a periodic resource usage pattern of VoIP traffic can be avoided. The signalling overhead can be reduced and the overall VoIP capacity increased.

There are already some proposals known to implement this technique: In [4] a concept called group scheduling is proposed. A group contains all MSs that are served by the same modulation and coding scheme (MCS). The position of a MS's resource within a group is fixed but the position of the group itself may change to fill resource holes. Thus, the number of resources allocations to be signalled is reduced. But changes in group memberships have to be signalled via an additional protocol.

The concept presented in [5] uses so called frame descriptor table (FDT) identifiers (IDs) to describe the current frame configuration. Each FDT ID references the resource usage pattern of a subset of MSs relative to the frame start. If parts of the resource allocations within a frame have the same pattern as in a previous one the base station (BS) only needs to signal the FDT ID that was assigned to the according resource usage pattern at its first use. Several FDT IDs can be transmitted per frame to signal non-overlapping resource usage patterns. A disadvantage of this concept is that the BS and all MSs need to provide a table mapping the FDT IDs to the according resource usage patterns. If the tables in the BS and MSs are not in sync this may result in harmful interference. Hence, a complex error handling protocol is necessary.

In [6] the suitability of dynamic and persistent scheduling for VoIP in 3rd Generation Partnership Project (3GPP) LTE networks is discussed. It concludes to use a hybrid approach where dynamic scheduling is the default and only when the signalling load becomes high, a part of the VoIP users is switched to semi-persistent scheduling. In the semi-persistent scheduling mode initial transmissions are scheduled persistently. Hybrid automatic repeat request (HARQ) retransmissions and silence descriptor (SID) frames that are transmitted to generate comfort noise are scheduled dynamically. Nevertheless, this method is also very complex as the allocation, reallocation and deallocation in persistent allocations have to be signalled. Furthermore, error-handling procedures need to be defined for the case that these information have not been or incorrectly received.

The scheduling scheme presented in [7] is similar to the previous one. Again, the allocation, reallocation and deallocation of persistent allocations are explicitly signalled by scheduling messages. The allocation is transmitted as MAC managemenet message together with the first VoIP protocol data unit (PDU). Revocation and changes of persistent allocations have to be signalled as extra messages. The combination of this protocol with automatic repeat request (ARQ) and HARQ mechanisms further increase the complexity of the proposed method.

In the current release of the Institue of Electrical and Electronics Engineers (IEEE) 802.16 standard [8] also a complex mechanism for the persistent allocation of resources is proposed. The advantages of reducing the signalling overhead are diminished through additional signalling necessary for allocation, reallocation and deallocation of persistent allocations. Additionally, sophisticated error handling procedures have to be established to avoid interference when signalling messages get lost.

In this paper, we propose a simple technique for persistent resource allocation (PRA) that increases the VoIP capacity of packet-switched systems considerably 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 WiMAX.

The paper is structured as follows: In Section 2 we introduce our concept for persistent resource allocations. Then, in Section 3 we present an analytical model for the evaluation of possible cell capacity gains in a single-cell scenario. The analytical results and their validation by event-driven simulation are shown in Section 4 and Section 5. The conclusions are summarized in Section 6.

2. CONCEPT

Voice traffic is characterized by constant VoIP packet sizes and inter arrival times (IATs) during talk spurts and no or very rare transmissions of comfort noise data during pauses. A frequently used voice codec is the adaptive multirate (AMR) codec [9] with a data rate of 12.2 kbs. It is compatible to the enhanced full rate (EFR) codec used in GSM

DOI: 10.1002/ett

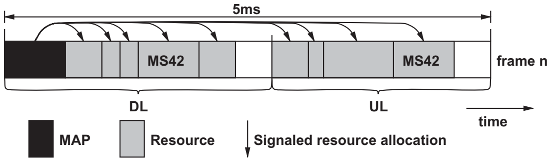

Figure 1. Without PRA all resource allocations need to be signalled at the frame start.

systems. The size of the voice packets is 244 bit and the IAT 20 ms during talk spurts. In silence periods only each 80 ms a comfort noise packet, also called SID, with 144 bit in size is generated. In our analytical investigations we ignore SID to keep the analysis simple.

In current frame-based packet-switched networks the necessary resources for voice services are periodically assigned to the MSs and signalled at the start of each frame in the so called MAP as indicated in Figure 1. The idea of PRAs is that resource allocations signalled once keep valid for 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 to set the validity of a PRA to a fix number of repetitions as depicted in Figure 2: the resource allocations for MS 42 are signalled in frame n (indicated by solid lines). They remain valid in frame $n + k$ up to frame $n + (m - 1)k$ (indicated by dashed lines). As the voice connection is ongoing and hence need further resources new resource allocations have to be signalled in frame $n + mk$. We call m time-to-live (TTL) as it limits the life time of the PRA. 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 shown in Figure 3. The resources of a frame envisaged for signalling resource allocations that are saved by applying the PRA scheme can be allocated to other VoIP connections. Thus, the number of VoIP connections carried per frame can be increased.

The two parameters m and k can be transmitted whenever a resource allocation is signalled. But, this results in

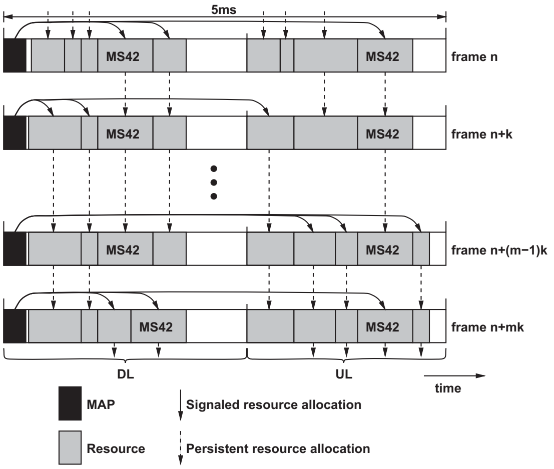

Figure 2. With PRA the average number of resource allocations that need to be signalled at frame start is considerably decreased reflected by a shorter MAP phase.

additional signalling overhead. As k is fixed for a certain VoIP codec and m can be assumed as a fixed value, too, we propose to negotiate both parameters once during service flow setup of the voice connection.

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 down-link (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 IEEE 802.16 recommend the use of unsolicited grant service (UGS). The BS periodically allocates DL and up-link (UL) resources for all service flows that use UGS 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

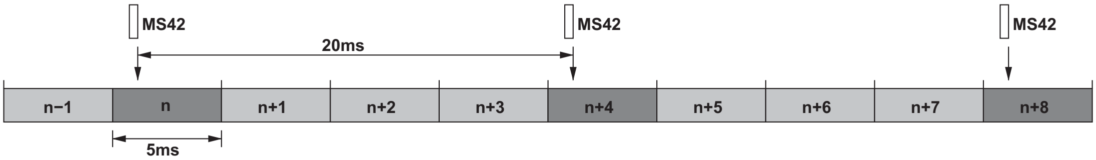

Figure 3. At a frame length of 5 ms and an IAT of 20 ms a single PRA is valid for each fourth frame.

DOI: 10.1002/ett

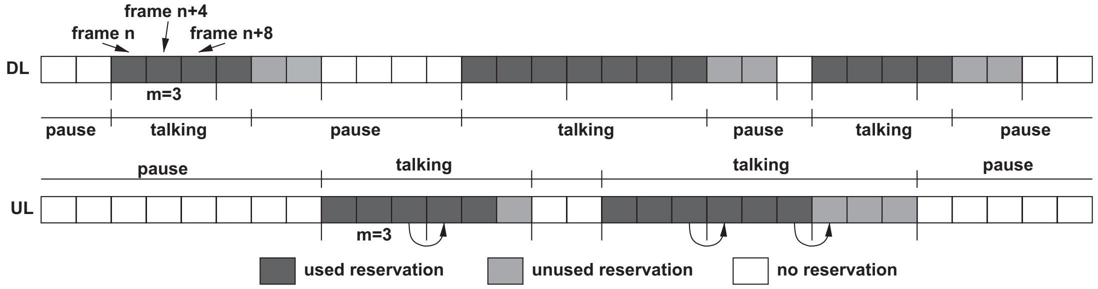

Figure 4. Resource usage of a single PRA enabled VoIP connection.

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. Instead, resources must only be allocated when needed.

If an MS intends to transmit VoIP data in UL direction it requests UL resources. If the BS supports the real-time polling service (rtPS) for that service flow this request can be transmitted when the MS is polled. Alternatively, the request can be transmitted contention based via bandwidth request slots in best effort (BE) service. Using rtPS yields to high signalling overhead as to all MSs that shall be polled sufficient resources have to be allocated per frame. Hence, usually BE service is preferred.

This leads to a problem for the UL direction in current packet-switched networks: if a MS gets assigned a requested 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 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 Figure 4, an exemplary resource usage of a single PRA enabled VoIP connection is shown. The TTL value is set to $m = 3$. In DL direction, the first resource of each PRA phase is always used. Otherwise no PRA would be signalled for that connection. The remaining resources of a valid PRA phase may be used. The same applies in UL direction when no valid PRA exists. To improve the statistical multiplexing of VoIP connections and to reduce the collision probability on the contention access 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 Figure 4. The differences in the usage patterns of the PRA phases have been considered for the analytical model presented in the following section.

Contention access for signalling resource requests and statistical multiplexing of VoIP connections lead to additional sources for packet errors: If several consecutive contention accesses of a MS result in collisions VoIP packets 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. The results in [10] indicate that the packet losses due to collisions during contention access or due to lack of resources are fairly rare if the number of contention slots is properly chosen and a call admission control mechanism limits the number of simultaneous calls to a reasonable degree.

3. ANALYTICAL MODEL

Figure 5 shows the Brady model [11] for bursty speech sources that is parameterized for VoIP speech. This model is required by the ITU-R for the evaluation of IMT-Advanced candidate systems [3]. The according model parameters are listed in Table 1.

As already mentioned, we evaluate the proposed technique exemplary for the IEEE 802.16 standard [8]. Hence, the following calculations base upon a typical parameter set for IEEE 802.16 systems. The size of a PHY PDU carrying a single VoIP packet is shown in Table 2. Payload

DOI: 10.1002/ett

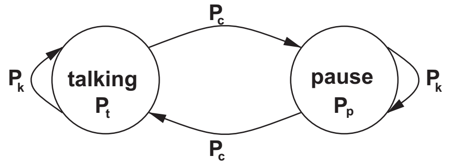

Figure 5. Brady model for VoIP traffic.

Table 1. VoIP model parameters defined by the ITU-R.

Model parameter	Value
Codec	RTP AMR 12.2
Encoder frame length	20 ms
$P_p = P_t$	0.5
P_c	0.01
$P_k = 1 - P_c$	0.99

header suppression (PHS) is an optional technique defined in the IEEE 802.16 standard to reduce the PHY PDU size by suppressing parts of higher layer protocol headers that do not change per packet. Only those parts of the headers that are continuously changing are transmitted within each PHY PDU. The constant parts of the headers are transmitted once at service flow setup and referenced by a so called payload header suppression index (PHSI) later on. For sake of completeness, Table 2 also shows the PHY PDU size for an optimal payload header compression technique.

Table 2. Size of a PHY PDU carrying a single VoIP packet.

	W/o PHS [bit]	With PHS [bit]	W/o headers [bit]
VoIP PDU size	244	244	244
RTP header	96	48	0
UDP header	64	0	0
IP header	160	0	0
MAC header	48	48	48
PHSI	—	8	—
Total PHY PDU size	612	348	292

Table 3. 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

Parts of the IEEE 802.16 MAC frame are reserved for special purposes such as ranging or the transmission of bandwidth requests. Therefore, these parts cannot be used for the transmission of resource allocations in the form of MAP information elements (IEs) or DL/UL traffic. In Table 3, the number of symbols for these MAC frame phases used in our analytical evaluation are listed. We assume a frame length of $t_{\text{frame}} = 5\,\text{ms}$ and a system bandwidth of 20 MHz. The evaluation is exemplary performed for the orthogonal frequency division multiplex (OFDM) PHY. Hence, a MAC frame consists of $N_{\text{frame}} = 360$ OFDM symbols. The number of symbols available for the transmission of MAP IEs and DL/UL data bursts then calculates to

$$N_{\text{ava}} = N_{\text{frame}} - N_{\text{pre}} - N_{\text{FCH}} - N_{\text{RTG}} - N_{\text{TTG}} - N_{\text{rang}} - N_{\text{bw}} \tag{1}$$

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 determined. 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 Reference [8] 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 Table 4. Additionally, the numbers of OFDM symbols necessary to transmit a PHY burst carrying a single VoIP packet are listed in the table. As the difference in the number of bit per PHY PDU when using PHS or optimal payload header compression is small, the PHY PDU sizes for both techniques map to the same amount of OFDM symbols for each PHY mode. Hence, in the following we consider only the cases where no header suppression technique or PHS is used. The results for optimal payload header compression are identical to the results for the PHS technique.

Table 4. Usage of MCSs and number of OFDM symbols per PHY burst in a free space scenario.

MCS	Coverage [%]	Bit per symbol	Symbols w/o PHS	Symbols w 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

Table 5. Average number of OFDM symbols per VoIP packet needed in the different frame phases.

	W/o SDMA		W SDMA	
	W/o PHS	W PHS	W/o PHS	W PHS
N_{avg}	4.55 sym	2.62 sym	1.41 sym	0.81 sym
DL-MAP	32 bit = 0.33 sym		48 bit = 0.5 sym	
UL-MAP	48 bit = 0.5 sym			
$N_{\text{UL-pre}}$	1 sym		0.31 sym	

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

$$N_{\text{sym,avg}} = \frac{\sum_{\text{MCS}} \frac{R_{\text{MCS}}}{100} \cdot N_{\text{MCS}}}{g_{\text{SDMA}}} \tag{2}$$

where R_{MCS} denotes the percentage of the coverage area by that MCS and N_{MCS} the number of symbols necessary to transmit the PHY burst. The parameter g_{SDMA} represents the capacity gain of the optional technique space division multiple access (SDMA). We assume that by using SDMA up to four MSs at different locations within the cell can be served the same time. The results in Reference [12] 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)).

In Table 5, the average numbers of symbols per PHY burst are listed for different combinations of the optional techniques PHS and SDMA. Furthermore, the sizes of a single DL ($N_{\text{DL-MAP}}$)/UL ($N_{\text{UL-MAP}}$) MAP IE are shown. The sizes are mapped to numbers of OFDM symbols assuming that the binary phase shift keying (BPSK)-1/2 PHY mode is used for their transmission. The optional SDMA mode cannot be applied during the transmission of the DL/UL MAP IEs as stated in [12]. The size of the UL burst preamble $N_{\text{UL-pre}}$ is shown in the last row of the table.

VoIP connections generate a symmetric traffic load. Hence, on average the same number of PHY bursts is transmitted in DL and UL direction. Therefore, the average PHY burst size N_{data} independent of the transmission direction calculates to

$$N_{\text{data}} = N_{\text{sym,avg}} + \frac{1}{2} N_{\text{UL-pre}} \tag{3}$$

The same applies to the MAP IE sizes:

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

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

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

As discussed in Section 2 statistical multiplexing of VoIP connections leads to further capacity gains. To estimate the VoIP capacity gain the resources of a valid PRA that are not used by a MS (see Figure 4) have to be considered as overhead as they cannot be reused by other MSs. This diminishes the overall capacity gain. In the following we derive the formula to calculate this overhead.

In Figure 6, all possible combinations of used/unused resource allocations of single PRAs are shown. In DL 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'). The same applies in UL direction for the first PRA after a 'pause' phase if no valid PRA exists. But, for succeeding PRAs already the first resource may not be used as 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 PRA usage patterns multiplied by the number of unused resources.

Equation (6) taken from [13] 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. With this value and the state change probabilities of the VoIP model the probability of a certain PRA resource usage pattern can be derived.

$$a(2^k + i) = a(2^k - i + 1) + 1 \text{ for } k \geqslant 0 \text{ and } 0 < i \leqslant 2^k \tag{6}$$

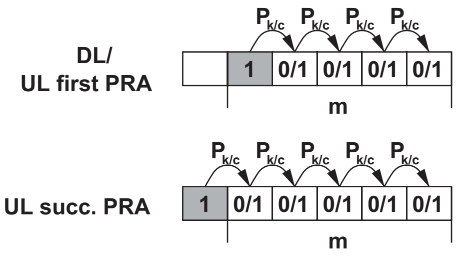

Figure 6. Resource usage patterns and transition probabilities.

DOI: 10.1002/ett

To calculate the number of zeros in a binary representation of an integer value Equation (7) is used. The formula is proposed in [14].

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

The combination of Equations (6) and (7) provides the average overhead $N_o(m)$:

$$N_o(m) = \sum_{i=0}^{2^{m-1}} \left(A_{i,m} b(i + 2^{m-1}) B_{i,m} \right) \tag{8}$$

with

$$A_{i,m} = P_c^{a(i+2^{m-1})}$$
$$B_{i,m} = P_k^{m-a(i+2^{m-1})-2}$$

As mentioned above, the PRA usage patterns in UL direction are different depending on the existence of a previous PRA. The Markov model shown in Figure 7 represents the three possible states of an MS transmitting in UL direction concerning PRAs: if it 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 calculated by Equations (9) and (10). Figure 9 shows how these equations are derived.

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

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

As shown in the upper part of Figure 8, $P_b(i)$ 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, subsequent 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 Figure 8 illustrates this. With probability $1 - P_b(i + 1)$ the MS does not request further UL resources and changes back to state 'idle'.

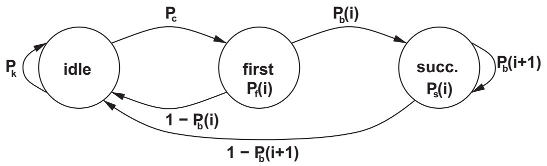

Figure 7. Markov chain representing the three possible states of an MS concerning PRAs.

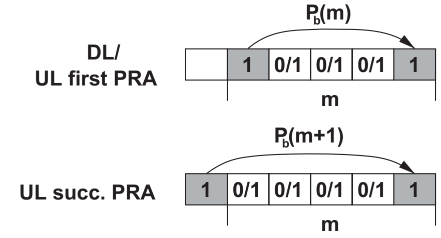

Figure 8. State change probabilities of the Markov model.

To calculate the average PRA overhead in UL direction only the ratios of the probabilities to be in state 'first' $P_f(i)$ and 'succeeding' $P_s(i)$ are relevant. These ratios are determined as follows:

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

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

Then, the mean overhead of PRAs $N_{\text{oh}}(m)$ independent of the transmission direction is

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

The overhead resulting from unused resources of a persistent allocation is increasing with m. By contrast, the overhead resulting from MAP signalling is decreasing for in-

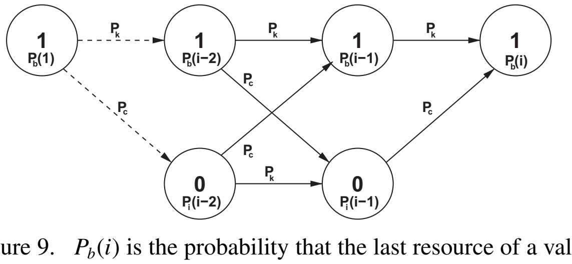

Figure 9. $P_b(i)$ is the probability that the last resource of a valid PRA is used.

creasing values of m. The total overhead considering both effects $N_{\text{oh,tot}}$ then calculates to

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

The total overhead per PRA $N_{\text{oh,tot}}$, the average number of OFDM symbols per PHY burst N_{data}, the number of OFDM symbols available per frame for transmission of MAP and DL/UL data N_{ava}, the IAT of VoIP packets t_{IAT} and the frame length t_{frame} can now be used to estimate the cell capacity $N_{S,\text{cap}}$ considering the voice activity factor $P_t = 0.5$:

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

The number of simultaneous calls $N_{S,\text{cap}}(m)$ is an average number. Hence, due to statistical effects it is possible that more UL or DL resources are needed than actually 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. But, in [10] it has been shown by simulation that already for a small number of guard resources the blocking probability is reduced considerably.

4. EVALUATION

Figure 10 shows the maximum number of concurrent calls over TTL for the UGS mode. As expected the number of concurrent calls increases with the TTL value but converges asymptotically towards a fixed limit that is different for each combination of the optional techniques PHS and SDMA. This results from the reciprocally proportional decrease of the MAP signalling overhead for increasing TTL values. The values for TTL = 1 correspond to the case that no PRAs are used at all. Therefore, the capacity gains through PRAs can be easily estimated: without using PHS and SDMA the capacity gain converges towards 8.2%, when using PHS towards 13.3%, when using SDMA towards 31.6% and when using both optional techniques towards 51.3%. When applying the optional techniques the gain increases as the ratio between the average size of a PHY data burst N_{data} and the average size of a MAP IE N_{MAP} decreases. The smaller this ratio is the higher is the gain.

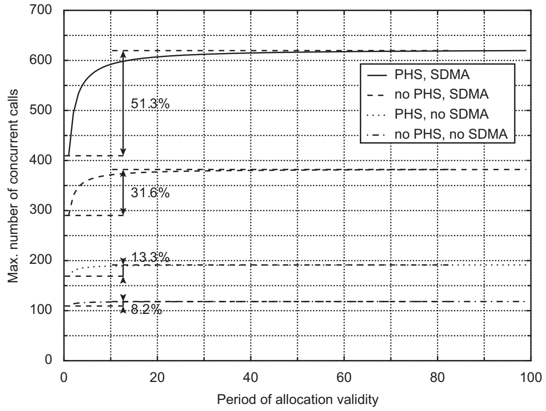

Figure 10. Max. number of concurrent calls carried over TTL for the UGS mode.

The graphs in Figure 11 show the total overhead $N_{\text{oh,tot}}$ over TTL for the case that persistent allocation is applied. These graphs reflect both effects that account for overhead: for values up to TTL = 10 the overhead resulting from the MAP IE signalling dominates. Hence, the total 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. Therefore, the overhead starts to increase, again. The optimal TTL value to minimize the total overhead is 10 and it is independent of the application of the optional techniques SDMA and PHS. We want to note that the number of VoIP packets is geometrically distributed. The state change probability from state 'talking' to state 'pause' is $P_c = 0.01$. Therefore, the mean number of VoIP packets per talk spurt is $E(X) = 1/P_c = 100$.

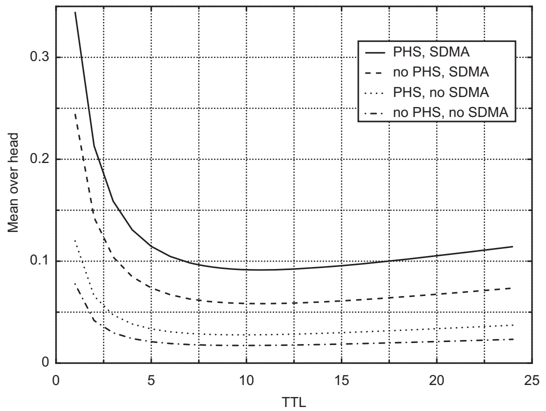

Figure 11. Mean overhead resulting from signalling of PRAs and unused resources of valid PRAs over TTL.

DOI: 10.1002/ett

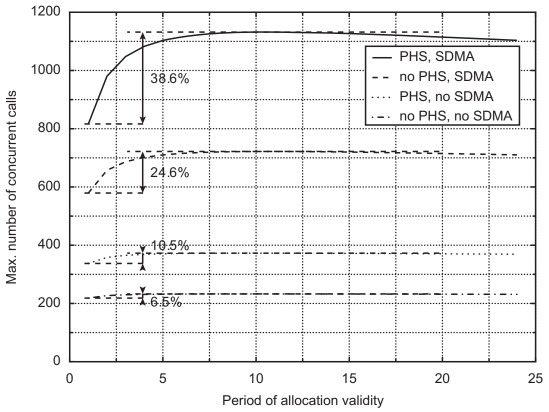

Figure 12. Max. number of simultaneous calls carried vs. TTL when using statistical multiplexing for VoIP connections.

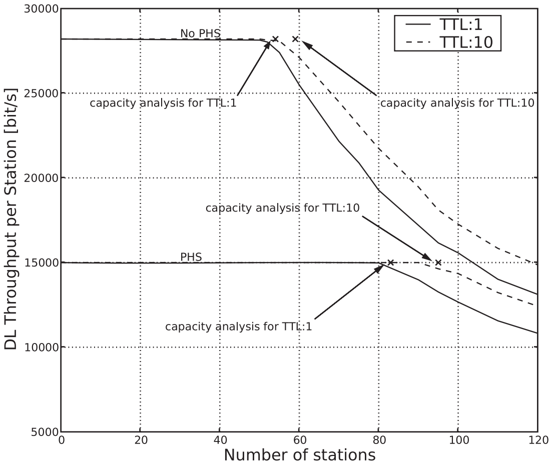

Figure 13. DL throughput per station over the number of stations for the UGS mode.

The corresponding VoIP capacity graphs are shown in Figure 12. Again, the maximum number of concurrent calls per cell are plotted over TTL. As expected the capacity gain for the optimal value TTL = 10 is maximal. Compared to TTL = 1 which again corresponds to the case that no PRAs are used at all the gain 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.

5. VALIDATION

The analytical model presented above has been validated by event-driven simulation. For this purpose the PRA concept has been implemented in the open Wireless Network Simulator (openWNS) [15]. The openWNS is currently under development at Communication Networks research group (ComNets).

The setup of the validation scenario meets the assumptions of the analytical model: A single cell with free space propagation conditions is simulated. The cell radius is set to the maximum distance that can be served by the most robust MCS. The WiMAX media access control (WiMAC), an openWNS module for the simulation of the IEEE 802.16 MAC sublayer is parameterized according to Table 3. Other openWNS modules implement upper layer protocols such as IP, user datagram protocol (UDP) and real-time protocol (RTP). The behaviour of the load generator complies with the requirements of the ITU-R model.

The graphs in Figure 13 show simulation results for the average DL throughput per MS at the MAC layer over the number of MSs for UGS without SDMA support. Again, the results for TTL = 1 (solid lines) correspond to the case that no PRAs are used at all. These are compared to results of scenarios with TTL = 10 (dashed lines). The value TTL = 10 has been chosen because it approximates the capacity limits for UGS already very well. Hence, the results for this value are plotted in the figure. The maximum capacity for the different operating modes is determined by the saturation points of the according graphs. To ease the validation, the according analytical results are plotted as crosses in the figure. As can be seen the absolute values for the cell capacity resulting from the simulations and the analytical evaluation match very well independent of the operating mode. The maximum error is below 10%. This shows that the analytical model of the PRA concept seems to be correct for the UGS mode.

6. CONCLUSIONS

The analytical results show that a considerable gain in VoIP capacity can be achieved through the PRA technique presented in this paper. Event-driven simulations confirm these results. The PRA technique is simple to implement and supersedes any error handling procedures required by other PRA techniques. Additionally, the technique enables the use of statistical multiplexing also for VoIP services as the number of resource requests for UL transmissions that are sent during the contention access phase are considerably reduced thus avoiding collisions and lack of resource allocations.

DOI: 10.1002/ett

ACKNOWLEDGEMENTS

The authors thank Professor Dr.-Ing. B. Walke for his fruitful support and friendly advice to this work. Furthermore, they acknowledge the nice cooperation with the ROCKET Consortium colleagues. This work has been performed in the framework of the FP7 project ROCKET IST-215282 STP, which is funded by the European Community (http://www.ict-rocket.eu).

REFERENCES

1. Sambale K, Klagges K. Increasing the VoIP capacity of WiMAX systems through persistent resource allocation. *15th European Wireless Conference, 2009*, Aalborg, Denmark, 2009; 308–313.
2. Walke BH. *Mobile Radio Networks—Networking, Protocols and Traffic Performance* (2nd edn). John Wiley & Sons, UK, 2001.
3. ITU-R M.2135: Guidelines for evaluation of radio interface technologies for IMT-Advanced. *ITU-R, Technical Report*, 2008.
4. Shrivastava S, Vannithamby R. Group scheduling for improving VoIP capacity in IEEE 802.16e networks. *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, 2009; 1–5.
5. Klein O, Einhaus M, Federlin A, Weiss E. Frame descriptor tables for minimized signaling overhead in beyond 3G MAC protocols. *Proceedings of Symposium on Trends in Communications (SympoTIC'06)*, 2006; 28–31.
6. Jiang D, Wang H, Malkamaki E, Tuomaala E. Principle and performance of semi-persistent scheduling for VoIP in LTE system. *Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing. WiCom 2007*, 2007; 2861–2864.
7. Suresh Kalyanasundaram SX, Bedekar A, Xu S, Xu H. Resource allocation scheme for 802.16m. *IEEE 802.16 Broadband Wireless Access Working Group, Technical Report*, C802.16m-07/258, 2007.
8. IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Broadband Wireless Access Systems. *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2001)*, 2009.
9. Mandatory Speech Codec Speech Processing Functions; AMR Speech Codec; General Description. *3rd Generation Partnership Project (3GPP), Technical Report* , TS 26.071, 2008.
10. Sambale K, Klagges K. On VoIP capacity gains in frame-based packet-switched wireless networks. *Leistungs-, Zuverlässigkeits- und Verlässlichkeitsbewertung von Kommunikationsnetzen und verteilten Systemen, MMBnet 2009 Workshop*, Hamburg, Germany, 2009.
11. Brady PT. A statistical analysis of on-off patterns in 16 conversations. *Bell System Technical Journal* 1968; **47**(1): 73–91.
12. Hoymann C. IEEE 802.16 metropolitan area network with SDMA enhancement. *Ph.D. Dissertation*, Aachen University, Lehrstuhl für Kommunikationsnetze, July 2008. Available online at: http://www.comnets.rwth-aachen.de
13. The On-Line Encyclopedia of Integer Sequences. AT&T Labs Research, 2009, Sequence number: A005811. Available online at: http://www.research.att.com/∼njas/sequences
14. The On-Line Encyclopedia of Integer Sequences. AT&T Labs Research, 2009, Sequence number: A080791. Available online at: http://www.research.att.com/∼njas/sequences
15. Bültmann D, Muehleisen M, Klagges K, Schinnenburg M, Walke B. openWNS - The simulation platform for IMT-Advanced systems. *European Transactions on Telecommunications*, European Wireless '09 Special Issue Paper, Wiley InterScience.

DOI: 10.1002/ett